Learning to organize digital information

Learning to organize digital information

Jaap Walhout

Open Universiteit
www.ou.nl

The research reported here was carried out at the

Open Universiteit
www.ou.nl

ISBN 978 94 6380 000 6

Cover design: Chris Peeters
Printing: Datawyse | Universitaire Pers Maastricht

UNIVERSITAIRE
PERS MAASTRICHT

Learning to organize digital information

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Open Universiteit
op gezag van de rector magnificus prof. mr. A. Oskamp
ten overstaan van een door het College voor promoties
ingestelde commissie in het openbaar te verdedigen

op vrijdag 26 oktober 2018 te Heerlen
om 16.00 uur precies

door

Jacob Herman Walhout
geboren op 24 juni 1971 te Borssele

Promotor

Prof. dr. F.L.J.M. Brand-Gruwel

Co-promotor

Dr. H.M. Jarodzka

Overige leden beoordelingscommissie

Prof. dr. T.A.J.M. van Gog, Universiteit Utrecht
Prof. dr. F.J.M. Huysmans, Universiteit van Amsterdam
Prof. dr. L. Kester, Universiteit Utrecht
Prof. dr. M.M. Specht, Open Universiteit
Dr. Y. Kammerer, Open Universiteit & Knowledge Media Research Center, Tübingen

TABLE OF CONTENTS

Chapter 1

General introduction

Whether we search for the longest river in Africa or gather information on how to deal with a disease, the Internet is the source to find any piece of information. Over the last decade the Internet has been transformed into a place where everyone can easily contribute, share and mix information, opinions, stories and creative expressions (e.g., Brand-Gruwel, Wopereis & Walraven, 2009; Martens, 2007). Consequently, nearly everybody uses the Internet to search for information (e.g., Jansen & Spink, 2006; Rouet, Ros, Goumi, Macedo-Rouet, & Dinet, 2011). This occurs in situations where the available knowledge is not sufficient to solve the problem at hand (Boekhorst, Kwast & Wevers, 2004; Case, 2007; Kuhlthau, 2004). Being able to search, find, evaluate, select, process, organize and present information to solve a problem are defined as information problem solving (IPS) skills (Brand-Gruwel, Wopereis, & Vermetten, 2005; Brand-Gruwel et al., 2009; Eisenberg & Berkowitz, 1990; Moore, 1995; Wolf, Brush, & Saye, 2003). In this thesis the main focus will be on the search aspect of IPS. Sometimes such information needs can be satisfied with a simple search query, such as when you want to know what the longest river of Africa is. However, a need for information can also arise from more complex, multifaceted situations (e.g., gathering information on how to deal with a disease). To satisfy such a complex information need, a more elaborate search process is required. This search process may span multiple sessions and can in some cases evolve in a recurring activity.

Besides deliberately searching for information, we are also unintentionally and constantly confronted with new information. While most of this information is irrelevant or does not have much prolonged value attached, some of the encountered information can elicit an already existing interest or may spark the development of a new interest. Such unintended discovery of useful information is often referred to as serendipity (Case, 2007). Whether the discovery of valuable information is the result of a thoughtful search or an unintentional encounter, when confronted with useful information we have to decide what to do with it (Whittaker, 2011). The main driver of this decision is whether or not we want to re-access this information at a later time.

Re-accessing is especially important when confronted with complex information problems, where found information has to be compared to previously found information (Rouet, 2003). It requires the collection of more than one source and in such complex tasks people need to re-access the information they found because they need to use it at a later time (Jones, 2007). This could be either at the end of one search session or at the end of a search process that spans multiple sessions. In this thesis several aspects of re-accessing information found on the Internet will be explored. Before elaborating on the re-access of information, the broader context of IPS has to be reviewed, because having a broad understanding of the concept of IPS is of importance to understand the issues around the re-access of information. The IPS-I model of Brand-Gruwel et al (2009) will be used to elaborate on this context. After reviewing the concept of IPS, the focus will be on the aspect of organizing information in the IPS-I model. The organization of information is of importance to eventually re-access information successfully.

Information Problem Solving as a complex cognitive skill

Whether it is for studying, personal interests or for work related tasks, people regularly have difficulties finding reliable information. There is no doubt that IPS skills are crucial in the digital society we are living in. However, research shows that many students are insufficiently able to solve information-based problems successfully. Studies of MaKinster, Beghetto, and Plucker (2002), Bilal (2000), Large and Beheshti (2000), McCrory Wallace, Wallace, Kupperman, Krajcik, and Soloway (2000), Van Deursen & Van Dijk (2008) and Walraven, Brand-Gruwel and Boshuizen (2008), reveal that students seldom know which search terms to use when searching the Internet, how to validly judge websites from the search results page, how to question the sources of the websites and the choice for opening a site is to a large extent guided by the title or summary of the site.

To design effective instruction and support these skills, it is necessary to have full understanding of the IPS process and its crucial aspects. Many models have been developed in the past to describe the IPS process (Fisher, Erdelez, & McKechnie, 2005). Although these models are very valuable and have much in common, they were developed in an era where the Internet was not the main source of information. There are some essential differences between searching in a library collection and searching for information on the Internet. The collection of libraries went through an, often rigorous, review process with the aim to filter out low quality information. On the other hand, on the Internet people can post information without any oversight on the quality of the information. This places an extra burden on the searcher as (s)he now also has to evaluate the reliability and trustworthiness of the information, a task formerly performed by librarians who were trained for it. Besides that, the fact that information became more instantly available also influenced the IPS process. These fundamental differences in the information environment required an update of the description of the IPS process. Building on prominent models like the search process model (Kuhlthau, 2004) and the Big6-model (Eisenberg & Berkowitz, 1990), Brand-Gruwel, Wopereis, and Walraven (2009) developed the IPS-I model to describe the IPS-process while using the Internet.

Based on empirical research regarding information problem solving by experts and novices, Brand-Gruwel et al (2009) presented a decomposition of the IPS process while using the Internet to search for information in the IPS-I model. This model distinguishes between main skills, regulatory skills and conditional skills. An overview of these skills is presented in Figure 1.

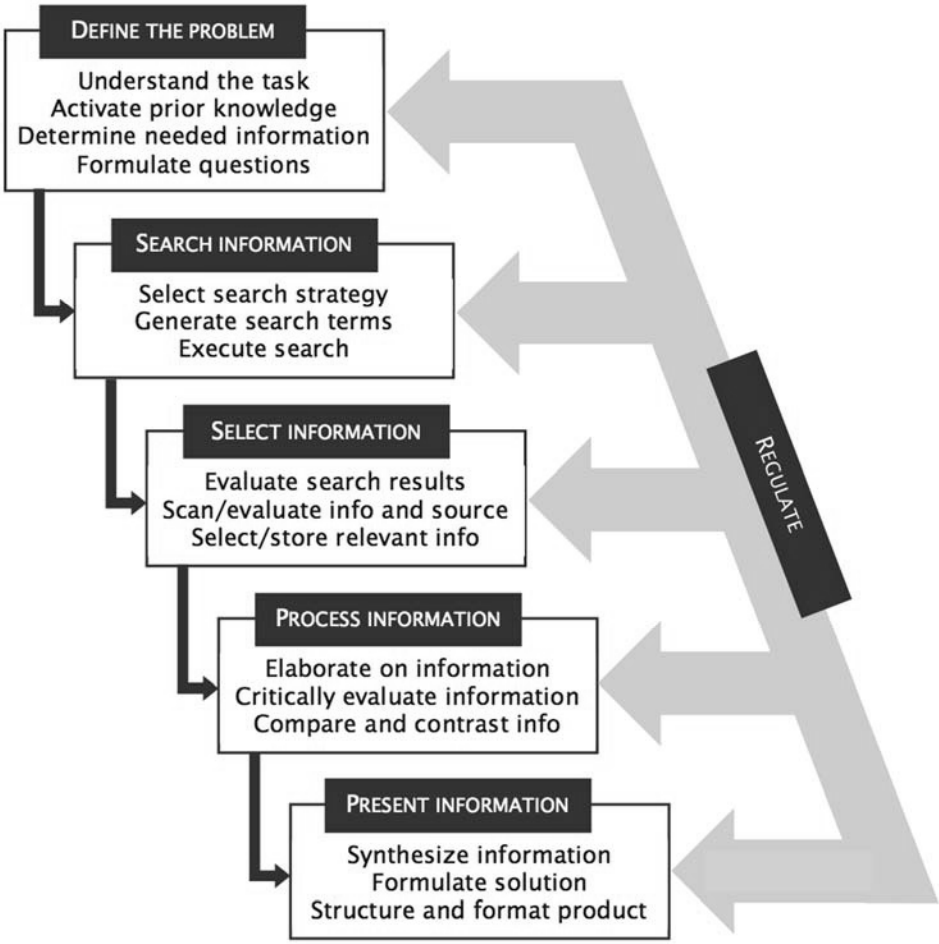

Figure 1: Overview of the information problem solving skill

In essence, this model depicts five main skills: defining a problem, searching for information, scanning information, processing information as well as organizing and presenting information. To solve an information problem successfully, students start by *defining the problem*. In this phase of the problem solving process, students need to activate the prior knowledge about the subject and determine what information is still missing. Based on this assessment, they formulate specific questions to guide them through their search process. After the problem definition, students have to come up with a *search strategy*, specify the search keywords and evaluate the list of search results. The most common used strategies are (1) searching with a search engine, (2) directly entering the Internet address in the browser and (3) browsing by just following links on webpages. When using the search strategy, students specify search keywords in the search engine. Subsequently, they evaluate their search results on relevance and reliability. Based on this evaluation, students open websites from the search result list to get a better idea about the contents of a webpage. During this *scanning of information,* students elaborate on the contents and try to link it to their prior knowledge, the problem description and other found information. When deemed useful they can store the information. Next, the collected information has to be *processed more deeply* to get a full understanding of the information and to connect the different found sources in a coherent manner with their prior knowledge. Then, they analyze the information and select and structure the useful parts. Finally, the information is *organized and presented*. The synthesis of the collected information should result in solving the information problem as defined in the problem definition.

Although the process seems to consist of five clear phases, in reality students tend to iterate between these phases constantly during the IPS-process. Moreover, for the successful application of these main skills, regulatory skills play an important role. Students should be able to orientate, monitor, steer and evaluate the IPS process constantly and be able to adapt the approach when needed. Furthermore, to regulate and iterate between the five main activities in the IPS process, students need to have decent level of proficiency in three conditional skills: Reading ability, evaluation skills and computer skills, which all support the IPS process. Before looking closer at strategies for re-accessing information, the role of organizing information in the IPS-I model will be explained.

Organizing information sources in the IPS-I model

In the IPS-I model, as presented in the previous paragraph, the activities of storing and organizing information play a role in several parts of the problem solving process (Brand-Gruwel & Wopereis, 2006; Brand-Gruwel et al, 2009). While scanning and processing information one has to store relevant information. The act of *storing* information is not a simple activity. Upon deciding to store information, one also has to decide where to store the information to enable easy access in the future. Simply keeping everything in one place is not a proper strategy. As Abrams et al (1998) already noted, to benefit from such a collection of information, it is necessary to somehow organize the found information. This is further illustrated by a survey of Symantec among lawyers (Symantec, 2010). Without exception, all respondents who filled out this survey indicated that difficulties in locating previously electronically stored information hampered legal processes in differing degrees of seriousness. As a result of these difficulties, lawyers experienced delays, were sanctioned by courts or even lost cases. This example illustrates perfectly that organizing information properly during the storing process is a necessity for future use. Unfortunately, the activities of storing and organizing information have received little attention in research thus far. At the end of the IPS process, students have to synthesize and present the information. Properly organizing information during the scanning and processing phases should be beneficial for the synthesis of the found information into an end product.

Strategies for re-access of information

As shown by the example above, organizing information is important when storing it. However, not all information that students search, needs to be stored for retrieval in the future. The need for storing and organizing information depends on the purpose of the search, the context of the information problem and how easy it is to re-find the information. First, people have different purposes when looking for information. A common used distinction in web search behavior is between navigational, transactional and informational search queries (Broder, 2002; Jansen, Booth & Spink, 2008; Rose & Levinson, 2004). A search query is considered a navigational query when the sole purpose of the query is to get to a specific webpage which the user already has in mind. The main goal of transactional search queries is to find resources (not being information). Examples of such transactional queries are downloading a report, watching a movie or using a currency converter. Lastly, informational search queries are mainly aimed at acquiring information about a certain topic. This could be to answer a question, to seek advice or to find out where a certain product or service can be bought. Storing and organizing information is not equally important for each of these purposes. It is probably most important in the case of gathering information about a specific subject, while it is less important during fact finding.

Second, to make things even more complex, the context is important as well. Broadly, three contextual environments can be distinguished: work, personal and educational. Education is nowadays more aimed at knowledge construction. As a result, students often get assignments where they have to search for information and have to process this information into an essay, a paper, or presentation. In the work context, organizing information and sources is necessary because a growing number of people can be characterized as ‘knowledge’ workers and therefore have to deal with large amounts of information. The study of Symantec (2010) showed that not storing and organizing information in a proper way is a barrier to performance in a work-related context. Organizing information can also have an added value in the personal context. Examples of this are hobbies where people collect information over longer periods of time or collecting information of personal interest, such as about a disease. The common feature of these three contexts is that information has to be stored for later use. However, the distinction between these contexts is not always clear and they can overlap.

Finally, the effort of searching again for the information has to be compared to the effort of storing and organizing the information immediately when first found. Relying on the strategy of searching again has the advantage that students do not have to think about how to store information. However, there are also some disadvantages to this strategy: (1) Earlier used search terms may be forgotten. When other terms are used to find a specific source again, it might happen that it will not show up in the search results. (2) The specific source is not available anymore. This could be due to a paywall or the source is not online anymore. (3) The search algorithm changed. As a result, the source will show up in another place in the search results or is even not included at all. (4) The specific source has been changed and is therefore less recognizable. Those disadvantages result in a decreased re-findability of information sources. The strategy of storing information overcomes these disadvantages mostly. However, the question of how interesting information sources can be stored arises. Aula, Jhaveri and Käki (2005) describe several strategies for saving information like saving documents as a file, printing, add bookmarks to favorites, email the url, or writing down the url's. It is obvious that these archiving methods do not guarantee a well-organized collection of information sources. Internet sources have to be stored in such a way that they can be easily found again. On the other hand, an experiment of Civan, Jones, Klasnja and Bruce (2008) shows that even organizing small sets of sources has an added value. In a small experiment, participants (students in higher education) were forced to store and organize 25 information sources with two existing email applications (Gmail and Hotmail). They could come up with their own classification scheme. Although not compared to a control group, these participants were able to re-access the information sources quite fast and were also able to remember details quite good, especially when using tags for organizing information.

In educational settings, in which the search queries are mostly intentional and often span multiple sessions, it is of particular interest how students re-access information they once found. As we have seen in the previous section, there are basically two strategies to re-access information: search again or store the information. In this thesis, the focus is on storing and organizing information. When storing information, people have to decide whether it is worthwhile to organize the information sources and how they want to do that (Abrams et al., 1998). Before the Internet, information was in books which were stored in libraries. Books are then stored on shelves. The main subject of a book was (and still is) therefore used to classify the book into a certain category (Wright, 2007). On the Internet, web directories were created to bring order in the available content on the Internet. The library way of organizing information was adopted, because people were used to such hierarchical storage systems, and it seemed natural to categorize information in classes and subclasses (Bush, 1945; Shirky, 2005; Wichowski, 2009). But problems arise when unusual or new pieces of information are difficult to classify, because they do not fit properly to the classification rules. Another problem emerges when an information object fits in more than one (sub)class (Oh & Belkin, 2014). The more information, the bigger this problem gets. Although hierarchical classification systems can bring order in a collection of physical objects, with the use of virtual information objects it is possible to organize the information with labels or 'tags'. In the tagging approach information object are not divided between folders but are given labels instead. Tagging allows for assigning multiple tags to an information object and also one tag to different information objects. Consequently, information can be attributed to different classes. With the emergence of social bookmarking, which allows people to define their own tags (folksonomy), users are given more control over organizing the contents of the web using own rules and categories.

Prior knowledge, task complexity and providing support

Prior knowledge provides people with internal cognitive schemata about a subject. Information found on the Internet can thus be compared to an existing knowledge structure. Consequently, more subject knowledge helps to organize newly found information (Bergman, et al., 2013a; Kalyuga et al., 2003; Kalyuga 2007; Kirschner et al., 2006; Rogers & Swan, 2004; Špiranec & Ivanjko, 2013). It is therefore much more difficult for novices to classify and organize new information. An additional barrier is the complexity of the task. More complex tasks are harder to solve and require more prior knowledge and better problem solving skills (Byström & Järvelin, 1995). In educational settings, students can be considered novices on most subjects. Their lack of prior knowledge hinders them to organize information in a meaningful way (Clark et al., 2012; White et al., 2009). Thus, novices need support to develop cognitive schemata (Van Merriënboer & Kirschner, 2012). An evident way of supporting novices in organizing information is providing them with a classification scheme (De Vries et al, 2008; Stadtler & Bromme, 2008). As De Vries et al (2008) showed, providing students with a detailed classification scheme leads to more correct answers compared to students who only received a small and shallow set of categories. Stadtler and Bromme (2008) also found that providing students support with a classification scheme improved their task performance. Notes were better organized and knowledge acquisition improved. These two studies indicate that providing students with a supporting classification scheme might lead to better results in solving IPS tasks.

Aim and outline of this thesis

The aim of this thesis is to investigate how novices, in particular students, can be supported in organizing found information during web search tasks. The main research question of these studies is, whether and when support with a classification scheme is beneficial and if so, whether it is better to support students with a hierarchical or a tagging system. Chapter 2 reports on the first study, which investigated the effect of task complexity on task performance. The effect of three search tasks of varying complexity, from simple to more complex, on search behavior and task performance was investigated. Chapter 3 presents a study investigating how well students were able to solve two information problems by navigating through information sources in a learning environment. In this environment the information was the same for all students (secondary education), but the type of navigation differed. They could either use a hierarchical menu or a tag cloud to navigate through the learning environment. Chapters 4 and 5 look more specifically at supporting students in organizing information while searching the Internet. Chapter 4 analyses the result of an exploratory study on the interaction between classification support and type of organizing system. Students solved an information problem by evaluating a list of search results. They had to organize the useful information with either a hierarchical or a tagging system and were supported with a classification scheme or had to make up their own, resulting in four conditions. The effect on bookmarking and tagging behavior and on the quality of the selected webpages was measured. In Chapter 5, support in organizing information was investigated in more depth. The main focus of this study was the effect of classification on how the support was processed and what effect the classification support had on tagging behavior and experienced task difficulty. Finally, Chapter 6 presents a synthesis of the main findings of the four studies described in this dissertation. Additionally, implications for educational practice and for future research are discussed.

Chapter 2

Effects of Task Complexity on Online Search Behavior of Adolescents

This chapter is based on:
Walhout, J., Oomen, P., Jarodzka, H., & Brand-Gruwel, S. (2017). Effects of task complexity on online search behavior of adolescents. *Journal of the Association for Information Science and Technology, 68*(6), 1449-1461. doi: 10.1002/asi.23782

Abstract

Evaluation of information during information problem solving processes already starts when trying to select the appropriate search result on a search engine results page (SERP). Up to now, research has mainly focused on the evaluation of webpages while the evaluation of SERPs received less attention. Furthermore, task complexity is often not taken into account. A within-subjects design was used to study the influence of task complexity on search query formulation, evaluation of search results and task performance. Three search tasks were used: a fact-finding, cause–effect, and a controversial topic task. To measure perceptual search processes, we used a combination of log files, eye-tracking data, answer forms, and think aloud protocols. Results reveal that an increase in task complexity results in more search queries and used keywords, more time to formulate search queries, and more considered search results on the SERPs. Furthermore, higher ranked search results were considered more often than lower ranked results. However, not all the results for the most complex task were in line with the expectations. These conflicting results can be explained by a lack of prior knowledge and the possible interference of prior attitudes.

Pupils often study a subject based on the information they can find online (Rouet, Ros, Goumi, Macedo-Rouet, & Dinet, 2011). For studying, as well as for everyday tasks, people experience problems in finding reliable information (Brand-Gruwel, Wopereis & Vermetten, 2005; Large & Beheshti (2000); Walraven, Brand-Gruwel & Boshuizen, 2008). The skills to find reliable information do not develop spontaneously and require instruction (Walraven et al, 2008). Surprisingly, the instruction of these information problem solving skills (IPS) receives relatively little attention in education. In their review study, Walraven et al. (2008) concluded that people in all age groups experience problems with IPS-skills. When searching for rather straightforward facts, formulating search queries and the shallow evaluation of search results is not a problem (e.g., searching for the height of the Eiffel tower). However, when people search for information about more complex issues, such as the causes of climate change, search queries can be formulated in different ways resulting in a diversity of information. For instance, one can search for the effect of humans on global warming but also search for causes of the greenhouse effect. The search results will be different in both cases. Furthermore, the chance that when working on complex tasks the top search results will represent a wrong or incomplete picture of reality, or even conflicting information is high (Kammerer & Gerjets, 2014). This is corroborated by Guan and Cutrell (2007) who found that irrespective of their relevance, people consider only the top search results. It is therefore important that pupils learn to evaluate and compare information from different sources.

The evaluation of information takes place in several parts of the information problem solving process (Brand-Gruwel, Wopereis & Vermetten, 2005; Brand-Gruwel, Wopereis & Walraven, 2009; Gerjets, Kammerer & Werner, 2011; Järvelin et al., 2015; Kuhlthau, 2004). The three consecutive phases of searching for information, scanning information and processing information all contain an evaluation sub-process. The first evaluation moment is when the search engine results page (SERP) is processed and a decision has to be made which search result to select. Deciding which webpage to open based on the title, the URL and the summary is not easy. The URL can, for instance, tell something about the organization behind the website. The second and third evaluation moments are related to judging the reliability and usability of the visited webpages itself and the information found on the pages. Questions like 'What is the organization behind the website?' 'What is the purpose of the website?' 'Who is the author?' 'Is the information up-do-date?' 'Is the information primary or secondary?' are of importance to judge the reliability. Up to now, research has mainly focused on the evaluation of webpages and documents (e.g., Walraven et al, 2008) while the evaluation of SERPs has received less attention. In this experiment we focused on the formulation of search queries and the evaluation of SERPs and their underlying processes. Moreover, we take the effect of tasks with different complexity levels into account.

The aim of our study is to get insight into the search phase of the information problem solving process with a focus on search query formulation and the evaluation of SERPs in relation to task complexity. To achieve this, we will integrate several process and task performance measures. Before we look into the processes of search query formulation and the evaluation of the SERPs, we will first discuss the concept of task complexity.

Classifying task complexity

In the literature the concepts of task complexity and task difficulty are both used interchangeably. As Gwidzka and Spence (2006) point out, these two concepts are not the same. Whereas task complexity is a more objective measure based on task characteristics, task difficulty describes the user perceptions of the task. Task difficulty is also often referred to as subjective task complexity. To avoid confusion, we will use task complexity in terms of an objective construct.

Experiments are often designed such that the requested information in search tasks can be found on one specific webpage (Puerta Melguizo et al., 2012), whereas in reality information is often scattered across webpages. This is especially the case in information-gathering search tasks where found information often has to be kept in mind or stored and new information has to be compared with this previously found information. As a result, these tasks are considered complex (Rouet, 2003).

Researchers have proposed different classifications of search tasks according to their complexity (Puerta Melguizo et al., 2012). Marchionini (1989) made a distinction between closed tasks and open-ended tasks, whereas Qiu (1993) calls them specific and general tasks. Walker and Janes (1999) and Kim and Allen (2002) distinguish know-item and subject search tasks. All four approaches more or less compare fact-finding tasks with information-gathering tasks.

Kellar, Watters and Shepherd (2007) proposed a more detailed distinction of four categories between types of search tasks. Besides fact-finding and information-gathering tasks, they also distinguish transactional and browsing tasks. Browsing refers to visiting web pages with no specific goal in mind, while transactional tasks refer to performing an online action such as webmail, banking or posting an update to Facebook or Twitter. A drawback of these studies (Kellar et al., 2007; Kim & Allen, 2002; Marchionini, 1989; Puerta Melguizo et al., 2012; Qui, 1993, Rouet, 2003; Walker & Janes, 1999) is that they focus more on classifying people's intent than task complexity. None of these studies provides a coherent methodology for classifying search tasks in all their subtle differences in complexity. As Wildemuth and Freund (2012) show in their review about exploratory search behaviors, the characteristics of search tasks differ along a variety of dimensions such as task goal, specificity, difficulty and (un)certainty. In a follow-up review, Wildemuth, Freund and Toms (2014) differentiated task complexity and task difficulty even more: open-ended and closed tasks, the degree of task structure, time aspects and the number of target items.

Although several studies describe the components of task complexity in detail (e.g., Järvelin et al., 2015; Vakkari, 1999; Wildemuth et al., 2014), it is a challenge to determine it objectively (Saastamoinen, Kumpulainen, & Järvelin, 2012). In particular, these studies do not use a clear method for determining differences in complexity between tasks in which also interactions between the used dimensions are taken into account. By using an additive model of several aspects, Mosenthal (1998) provides a flexible framework for determining such differences in task complexity and incorporates most of the aspects mentioned by Wildemuth et al. (2014) or Kelly, Arguello, Edwards and Wu (2015). In this framework, task complexity is evaluated along three different dimensions: type of information requested, type of match, and plausibility of distractors. *Type of information requested* describes to which extent the to-be-found information is concrete (i.e. a fact, place, or person) or abstract (i.e. a reason, evidence, or condition). Finding concrete information is easier than abstract information. *Type of match* covers the resemblance between the information in the question and the information in the source, and the level of inference between those two. The type of match scale comprises a set of variables, based on which the final value is determined. Besides the type of task (locate, cycle, integrate, or generate), the outcome value is also influenced by whether information has to be compared or contrasted; the number of phrases in the question; how many items an answer contains; whether the match is literal, synonymous or has to be inferred for given information; and finally whether no, low or high inference is needed for the requested information. The *plausibility of distractors* scale describes the degree of distracting information (i.e., similarity to the required information). In this study we used Mosenthal's framework to determine the complexity of tasks.

Search behavior: search query formulation

The actual searching for information on the Internet often starts with the formulation of search queries. People use three keywords on average to formulate search queries (Jansen, Spink & Koshman, 2007). Belkin et al. (2003) conclude that using longer search queries results in increased search effectiveness. However, Aula and Nordhausen (2006) found that longer search queries do not result in shorter task completion time.

Search keywords are straightforward for fact-finding tasks as they can be deduced from the task directly. Complex search tasks require more elaborations on the search keywords (Borlund & Dreier, 2014). The choice of keywords depends on the topic, the familiarity with that topic and, in case of controversial topics, on the attitude. For instance, if you believe that we humans are the major cause of climate change, you will search for information that confirms your ideas and select keywords accordingly, when searching for its causes (Veldhorst, Frerejean, Van Strien, & Brand-Gruwel, submitted).

Additionally, for accomplishing complex tasks people use more search queries, more search keywords and adapt the search query more often (Aula, Khan & Guan, 2010; Barsky & Bar-Ilan, 2012; Borlund & Dreier, 2014; Kelly et al, 2015; Liu, Gwizdka, Liu, & Belkin, 2010; Saastamoinen et al., 2012; Singer, Norbisrath, & Lewandowski, 2012a; Toms et al., 2008). People are also more likely to adapt their search queries rather than going to the second results page or further (Jansen & Spink, 2006; Choi, 2010). Furthermore, the adoption of query operators (e.g., double quotes, plus- or minus sign, or the 'site:' operator) remains quite low (Aula & Nordhausen, 2006; Jansen et al, 2007).

Evaluating search engine results: Processing of SERPs

The first evaluation within the IPS process takes place on the SERP. The studies that took the evaluation of SERPs into account show that people highly trust the top-ranked search results (Guan & Cutrell, 2007; Hargittai et al., 2010; Huvila, 2013; Pan et al., 2007; Westerwick, 2013). Most people also select these results (e.g., Höchstötter & Lewandowski, 2009; Salmerón, Kammerer & García-Carrión, 2013) without deep considerations (Kammerer & Gerjets, 2014; Lorenzen, 2002; Salmerón & Kammerer, 2012; Walraven et al., 2009). However, this is an error-prone approach for several reasons. First, people assume that search engines are neutral in presenting the results. However, as Machill, Neuberger, Schweiger and Wirth (2004) state, the near monopoly of Google is problematic because its criteria to order the results are not clear. Secondly, certain websites, such as Wikipedia, are shown more often in the first search results (Höchstötter & Lewandowski, 2009; Finley, 2015). Third, many websites try to get a higher rank with the use of search engine optimization (Höchstötter & Lewandowski, 2009; Machill et al., 2004; Lewandowski, 2012). Fourth, despite the enormous amount of information sources, people often restrict themselves to only a few, mostly those used before (Head & Eisenberg, 2011; Purcell et al., 2012). Fifth, the quality of the presented search results differs between search engines (Lewandowski, 2015). Sixth, the presented text captions in the search results are often biased towards the search query (Tombros & Sanderson, 1998; White, Jose & Ruthven, 2003) and are not always a good prediction of the underlying webpage relevance (Bailey et al., 2010).

In previous research on the evaluation of SERPs, often one or more of the following process measures are used: eye movements, activity logs, and/or verbal data. Eye movement data revealed that in complex tasks people viewed more SERPs and more unique SERPs (Liu et al., 2010). Log data showed (Höchstötter & Lewandowski, 2009; Jansen & Spink, 2006; Silverstein, Henzinger, Marais, & Moricz, 1998) that most people do not visit more than one SERP and often do not scroll to the bottom (Höchstötter & Lewandowski, 2009). As a result, the top search results are viewed more often than the results at the bottom (Granka, Joachims & Gay, 2004; Guan & Cutrell, 2007; Höchstötter & Lewandowski, 2009; Pan et al, 2007). Kammerer and Gerjets (2012) compared a standard Google-like SERP with a manipulated SERP where the search results were presented in a table. They demonstrated that the way of presenting the search results influences viewing and selection behavior on SERPs. However, this is partly contrasted by Kammerer and Beinhauer (2010). They compared a regular list interface with a tabular interface and a grid interface. They found no significant difference between types of interface.

When analyzing activity logs, we found that students used less time to evaluate overview pages in a more complex task (information-gathering) than in a fact-finding task (Walhout et al., 2015). This is in line with the findings of Singer et al. (2012a). They found that for complex tasks people spent more time on the SERPs, selected more search results and needed more time to complete the tasks.

With regard to verbal data, Gerjets, Kammerer and Werner (2011) tested whether explicit evaluation instructions influence the evaluation processes of SERPs compared with spontaneously produced evaluations. Explicit evaluation instructions lead to significantly more verbalized thoughts. This provides a clear guideline on how to derive verbalizations of thoughts in research. Analysis of verbalized thoughts by Walraven, Brand-Gruwel and Boshuizen (2009) revealed that when evaluating the search results, students emphasize the rank in the list and the title/summary of the search result as the most important evaluation criteria. This study also showed that students know more evaluation criteria than used when asked to verbalize their thoughts. When explicitly asked to include aspects of relevancy, the quality of peoples' judgements increases significantly (Kim, Kazai & Zitouni, 2013).

Task performance

Although the reviewed studies provided interesting insights into behavioral aspects with regard to search query formulation and the evaluation of SERPs, most of them have not taken task performance into account. Still, several related studies provide hints of what to expect. Byström and Järvelin (1995) showed, for example, that with increased task complexity, the success rate of search tasks decreases, the need for domain and problem solving information increases and that the number of sources increases. Vakkari (1999) concluded that task complexity is a crucial factor in determining task performance. Furthermore, Singer, Pruulmann-Vengerfeldt, Norbisrath, and Lewandowski (2012b) found that previous experience has a weak positive relation with task performance. However, this relationship is even less evident for complex search tasks than for simple search tasks. Besides the quality of the answer, time-based measures are also a valuable indicator for task performance. Previous studies conclude that successful searchers need less time to complete the search task (Jansen & Spink, 2006; Singer et al., 2012a). We also found in an earlier study (Walhout et al., 2015) that simple fact-finding tasks elicited more active browsing behavior than complex information gathering tasks: participants visited more pages and spent less time on viewing pages. This is due to more elaborate reading and evaluation of information involved in information-gathering tasks. This is in line with the findings of other studies (Liu et al., 2010; Singer et al., 2012a). Furthermore, several studies showed that more complex tasks require more time on task (Borlund & Dreier, 2014; Brennan, Kelly & Arguello, 2014; Liu et al., 2010; Kelly et al, 2015) as well as for more specific time-based measures like search time, time on SERPs and reading time (Singer et al., 2012a).

THIS STUDY

While the presented literature gives valuable insights in how people formulate search queries and evaluate online searches, not much is known about how these processes differ between tasks with different levels of complexity and what the effects are on task performance. This is the aim of the present study. Detailed time-locked process-tracing methods such as eye tracking, logging data, and thinking aloud protocols were combined to get more insight into the evaluation behavior of students with respect to task complexity.

The central research question for this study is: How does task complexity influence behavior and task performance in online search? In particular: 1) Does task complexity have an effect on search query formulation? 2) Does task complexity have an effect on the evaluation of the SERPs? 3) Does task complexity have an effect on task performance?

Concerning differences in *search query formulation* the number of search queries and the number of keywords used per query are often used as measurements (e.g.: Barsky & Bar-Ilan, 2012; Borlund & Dreier, 2014; Liu et al., 2010; Singer et al., 2012a; Toms et al., 2008). Besides these, we measure the time to formulate a search query and the degree in which the used keywords match with the words in the task descriptions. A more complex search task will most likely lead to more time to think about how to formulate the search query. Furthermore, using words that appear in the task description can often be used to solve fact-finding tasks. This is less often the case for more complex tasks.

With respect to the *evaluation of search engine results pages*, the first important measurement is how much time the participants spend on the SERPs as this gives an indication about how much effort they put into evaluating the search results. Another useful measurement is how much time the participants need before they make their first selection and the number of websites they select. Ideally people consider all search results on a SERP before making a selection; but as earlier research shows, the top search results are viewed more often that the ones at the bottom of a SERP (Höchstötter & Lewandowski, 2009; Pan et al, 2007). We do not know yet whether and how task complexity affects the viewing behavior on SERPs. Finally, we will also incorporate verbalized thoughts to get a better understanding of the evaluation of SERPs.

With regard to *task performance* not only the quality of the answer was taken into account, but also time on task and the time used to formulate an answer were used as performance measurements.

METHOD

Participants and design

A within-subjects design was used with task complexity as the within-subjects factor. The dependent variables were search query formulation, evaluation of search results, and task performance.

Information retrieval studies often use university students or employees as subjects and the use of different populations is therefore encouraged (Kelly et al., 2015). In this experiment 15 students (8 male and 7 female; mean age 14.5 years, SD = 0.92) from the highest level of secondary education in the south part of The Netherlands participated. All participants had experience with searching on the Internet, but did not receive any information literacy training prior to the experiment. Participation was voluntarily and was rewarded with a gift certificate of 15 euro. This study adheres to local ethical requirements.

Material and apparatus

Tasks. During the experiment, participants solved three information problems with the Google® search engine: a fact-finding task, a cause-effect task, and a controversial topic task. During the tasks they could search the Internet without any restrictions and without time limit. Each task was answered with a short text. In designing experimental tasks, a trade-off has to be made between experimental control and realism (Borlund, 2000). A simulated task needs to have a content to which the participants can relate, has to be of interest to them and should provide enough imaginative context (Borlund, 2000; Bell & Ruthven, 2004). The tasks were constructed in cooperation with teachers to ensure that the task content and workload were comparable to what is common in education. Table 1 gives an overview of the three task descriptions.

Table 1 Tasks

Fact-finding	What is the habitat of the Damara zebra?
Cause–effect	Why does the road sometimes look like a mirror in the distance?
Controversial topic	Does radiation from mobile phones have consequences? If so, what are these?

Task complexity of the three tasks was assessed with Mosenthal's framework of prose task characteristics (Mosenthal, 1998). Task complexity was judged along the three different dimensions of this framework: type of information requested, type of match, and plausibility of distractors. Table 2 summarizes the scores on each scale for the three tasks.

Table 2 Task complexity scores

	Type of information requested	Type of match	Plausibility of disctractors	Total score
Fact-finding task	1	2	1	4
Cause-effect task	4	5	2	11
Controversial topic task	4	9	4	17

The scale for *type of information requested* ranges from one (easiest) to five (most difficult). Because the requested information is very concrete in the fact-finding task, a score of one was given. The other two tasks received a score of four because they required the identification of a cause-effect relationship or evidence. The scale *type of match* ranges from one to twenty and is composed of a set of component variables: type of task; whether the information has to be compared or contrasted, number of phrases in the question; whether the match is literal, synonymous or has to be inferred for given information; and finally whether no, low or high inference is need for the requested information. For the fact-finding task a score of two was given because the participants had to locate the desired information between several possible texts. No extra points were given for the other components. For the cause-effect task three extra points were assigned because the match requires low text-based inference and the number of features of the required answer was unspecified. Four additional points were given for the controversial topic task as the match requires more text-based inference, requires more prior knowledge and the plausibility of distractors is higher.

Equipment. Eye movements and logging data were recorded with a remote Tobii 1750 eye tracker with temporal resolution of 50 Hz (Tobii, 2003), which is integrated with a PC screen, and is operated with Studio software (Tobii, 2007) from the stimulus PC (see www.tobii.com). The screen capture recording mode was used so that not only the eye movements, but the entire task performance process (including possible mouse- and keyboard operations and visited webpages) was captured. Data processing and data analysis was done with the R language for statistical computing (R Core Team, 2017).

Think aloud instructions. Even though Gerjets et al (2011) found a difference in verbalized thoughts between explicit evaluation instructions and neutral verbalization instructions, they did not find the expected differences in eye movement measures and in mouse clicks for the evaluation of SERPs. However, as they argue, explicit evaluation instructions might interfere with behavioral aspects as viewing and clicking behavior. To avoid such a possible interference, in line with Gerjets et al (2011) a standard neutral thinking aloud instruction was used. The instruction given to the participants was:

> Please think aloud during your web searches, verbalize everything you think about. You don't have to tell what you do, for example where you click, but only what you think and why you do things.

When the participants stopped verbalizing their thoughts during the execution of the tasks, the experimenter reminded them. The verbalized thoughts were recorded with Audacity audio recording software using a standard microphone attached to a separate laptop.

Measurements

Search query formulation. Search query formulation was measured with four variables: 1) number of search queries, 2) number of keywords per query, 3) time to formulate first query and 4) the degree in which the used keywords match with the task descriptions. The *number of search queries* was counted as the number of unique search queries. This number also represents the number of visited SERPs. When a search query is entered in the search engine, a SERP immediately appears. Consequently, the number of visited SERPs is equal to the number of search queries. Revisits to SERPs (i.e. after visiting a webpage) were not included. The *number of keywords* was calculated by dividing the total number of keywords used by the number of search queries. The *time to formulate the first query* was calculated as the difference between the starting timestamp and the timestamp at which the first search results page was accessed. The *degree in which the used keywords match* with the task descriptions was calculated as a relative measure ranging from zero (none of the used keywords matches words from the task descriptions) to one (all used keywords are in the task descriptions). To take into account spelling errors and minor word variations, a Levenshtein distance calculation with a maximum allowed distance of one was used to determine a match between a used keyword and the words from the task description.

Evaluating search results. The evaluation behavior of the search results was measured with several variables. From the logging data the following metrics were calculated: 1) total time spent on SERPs, 2) time to first selected search result and 3) number of selected search results (i.e. webpages). From the eye-tracking data the following metrics were calculated: 1) position of the selected search results on the SERP, 2) number of viewed search results, 3) number of viewed search results per query and 4) viewed search results by position on the SERP. The difference between a selected and a viewed search result is that a selected search result is both viewed as well as clicked on, while a viewed search result is only viewed. The *eye tracking parameters* were calculated for all visited search pages. All parameters were assigned to certain elements of these pages, so-called *areas of interest* (AOIs). Each search result on each visited search engine results page was defined as an AOI for each participant, see Figure 1. The AOI's of all participants were grouped in search result position groups. For example: the AOI's of each first search result were pooled into one group. For each search result position AOI the fixation duration and click count were determined. All analyses were performed with Tobii Studio software version 2.2.4 (2007). Before further analyses were executed, the raw data was filtered with the Tobii ClearView fixation filter. Based on visual inspection of the data, a fixation definition of 30 pixels and 100 milliseconds was chosen (cf. Hegarty & Just, 1993; Loftus, 1981; Cutrell & Guan, 2007; Gerjets, Kammerer & Werner, 2011).

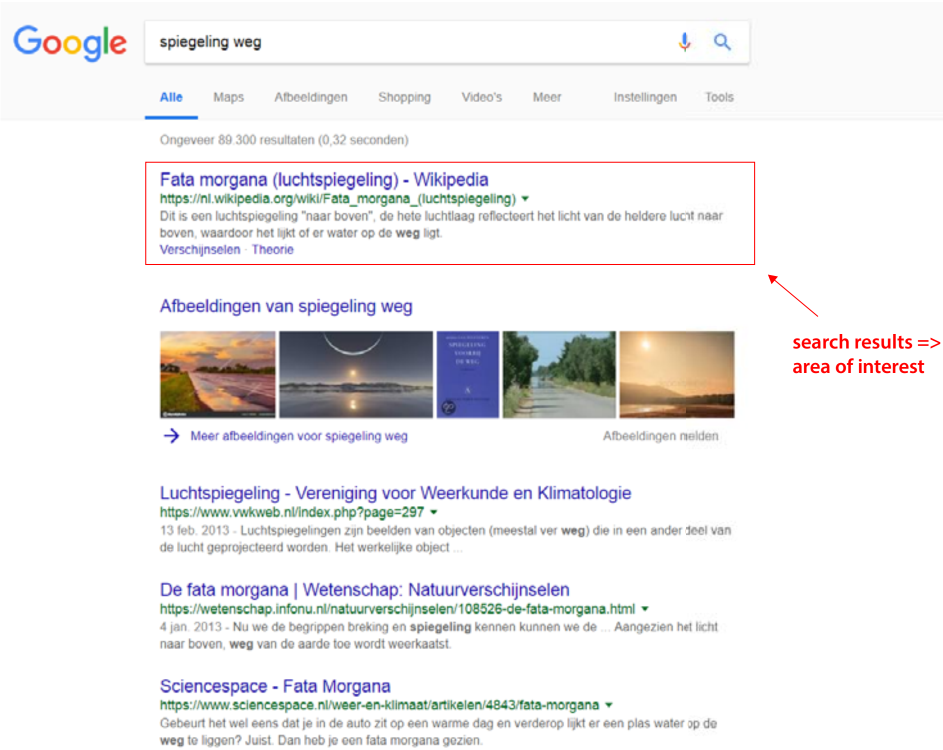

Figure 1: Area of interest on SERP

Because the search process is the main focus of this article, only the verbalized thoughts with regard to the evaluation of the SERPs were transcribed and analyzed with a coding scheme (based on Walraven, et al, 2009). The verbalized thoughts were transcribed and coded into five categories: evaluative thoughts which relate to 1) vague evaluation (I think this is a good website, or I think I can use this website), 2) the kind of webpage, 3) the web address, 4) the rank in the list of search results or 5) reputation of the organization behind the website. Two raters individually scored all protocols. The interrater reliability, expressed in Cohen's Kappa, was 0.83.

Task performance. Task performance was measured using three variables: answer quality, time on task, and answering time. *Answer quality* was measured as the correctness of the answers to the questions in the tasks. Answer quality was calculated as the number of correctly mentioned aspects in the answer divided by the maximum number of correct aspects, thus resulting in a score between zero and one. For the fact-finding task the correct answer had to contain the elements '*savannah*', '*east*' and '*Africa*'. The correct answer for the cause-effect task had to contain the concepts '*warm air*', '*cold air*', '*just above the asphalt / road surface*', '*light refraction*' and '*air layers*'. For the controversial topic task, the answer had to contain the elements '*scientific research*', '*no (health) effect*', '*no evidence*' and '*low level of radiation*'. *Time on task* was measured in seconds and indicated how long it took participants to complete the task. *Answering time* was measured in seconds and gives an indication about the difficulty of answering as well as commitment to the task.

Procedure

Prior to the experiment participants received an introduction to the study. Their parents were also informed about the study beforehand. The experiment was conducted in individual sessions. Each session started with a detailed explanation of the procedure. For each participant, demographic data (i.e., age, gender) was collected. Next, they were introduced to the eye tracking equipment. The eye tracker was adjusted to the individual features of each participant by calibrating the system with a 9-point calibration. There was no time constraint for the tasks. The duration of the longest session for all three tasks was approximately 50 minutes.

RESULTS

The results of this study are presented in three parts: search query formulation, evaluation of SERPs, and task performance. All relevant means and standard deviations are summarized in Table 3. A repeated measure ANOVA was used to analyze the data. The post hoc analyses were done with pairwise comparisons with Bonferroni adjustment. Because some measurement variables were not normally distributed and also transformations did not result in normally distributed data, Friedman's ANOVA with post hoc tests were used to analyze the data.

Table 3 Means and standard deviations

	Fact-finding (FF)	Cause-effect (CE)	Controversial topic (CT)	p-value	differences
N	15	15	15		
Task Performance					
Answer quality	0.67 (0.18)	0.51 (0.33)	0.07 (0.15)	< 0.01	FF > CT, CE > CT
Time on task	170.78 (138.89)	315.37 (148.98)	281.30 (143.20)	< 0.01	FF < CE, FF < CT
Time to formulate an answer	16.05 (8.43)	43.96 (31.51)	38.18 (29.44)	< 0.01	FF < CE, FF < CT
Search query formulation					
Search queries	2.53 (1.92)	3.73 (2.05)	1.73 (1.10)	< 0.01	CE > CT
Words per query	3.03 (0.67)	5.00 (2.39)	3.98 (1.06)	< 0.01	FF < CE, FF < CT
Time to formulate first query	9.06 (2.95)	22.92 (16.18)	13.24 (7.02)	< 0.01	CE > FF, CE > CT
Degree of matching keywords	0.79 (0.21)	0.62 (0.23)	0.92 (0.17)	< 0.01	CE < CT
Evaluation of SERPs					
Time spent on search pages	42.85 (16.31)	47.83 (13.20)	26.46 (16.72)	< 0.01	CT < FF, CT < CE
Time to first selected webpage	15.35 (17.67)	24.45 (21.91)	10.65 (9.26)	< 0.01	CT < CE
Selected webpages	3.13 (2.56)	3.47 (2.17)	4.33 (3.87)	> 0.05	
Viewed search results	17.46 (20.45)	24.93 (21.73)	7.93 (2.86)	< 0.01	CE > CT
Viewed search results per query	6.22 (2.54)	5.90 (2.36)	4.93 (5.50)	> 0.05	

Differences in search query formulation

To answer the question "Does task complexity have an effect on search query formulation?" it was analyzed whether the number of search queries, number of words per search query, time to formulate the first query, and match of keywords with the task description differed between the three tasks. In the cause-effect task, the participants needed a significantly higher *number of search queries* in comparison to the controversial topic task ($\chi^2(2) = 11.49$, $p < 0.01$). The other two post hoc comparisons showed no significant difference. The *number of keywords per search query* was significantly lower in the fact-finding task compared to the cause-effect task and the controversial topic task ($\chi^2(2) = 12.81$, $p < 0.01$). In the cause-effect task participants needed considerably more *time to formulate the first search* query than in the other two tasks ($\chi^2(2) = 16.53$, $p < 0.01$). There was no significant difference between the fact-finding task and the conflicting topic task. The *number of matching keywords with the task description* was significantly lower in the cause-effect task compared to the controversial topic task ($\chi^2(2) = 10.16$, $p < 0.01$). The other two post hoc comparisons showed no significant difference.

Differences in evaluation of search results

The analyses conducted regarding search behavior should give an answer for the question: Does task complexity have an effect on the evaluation of the SERPs? The operationalization of evaluation behavior is based on logging data as well as eye-tracking data. The following measurements were calculated: total time spent on SERP, time to first selected search result, number of selected search results, position of the selected search results on the SERP, number of viewed search results, number of viewed search results per query and viewed search results by position on the SERP.

In order to be able to do a meaningful analysis with regard to the *total time spent on SERPs*, the absolute time variable was transformed into a relative variable. The analysis showed that the participants spent significantly less time on the SERPs in the controversial topic task ($F(2, 28) = 16.16$, $p < 0.01$, $\eta^2 = 0.271$) compared to the other tasks. There was no significant difference between the fact-finding task and the cause-effect task. In the controversial topic task the participants *selected the first* search result (webpage) more quickly than in the cause-effect task ($F(2, 28) = 6.01$, $p < 0.01$, $\eta^2 = 0.135$). Although the other two post hoc comparisons showed no significant difference, the difference between the fact-finding task and the cause-effect task had a p-value below 0.10 and could therefore be considered a trend. Although the *number of selected webpages* was the highest for the controversial topic task, the difference was not significant ($\chi^2(2) = 1.17$, $p < 0.05$). The *positions of the selected webpages* on the SERP are visualized in Figure 2. In the fact-finding task and in the cause-effect task, the participants show a tendency to select the first search result most often. In the controversial topic task this preference is less obvious. Table 4 summarizes the skewness and kurtosis values for the distribution of the selected search results by position on the SERP. These values show that the concentration of search result selection is most concentrated to the top for the fact-finding task and least for the controversial topic task.

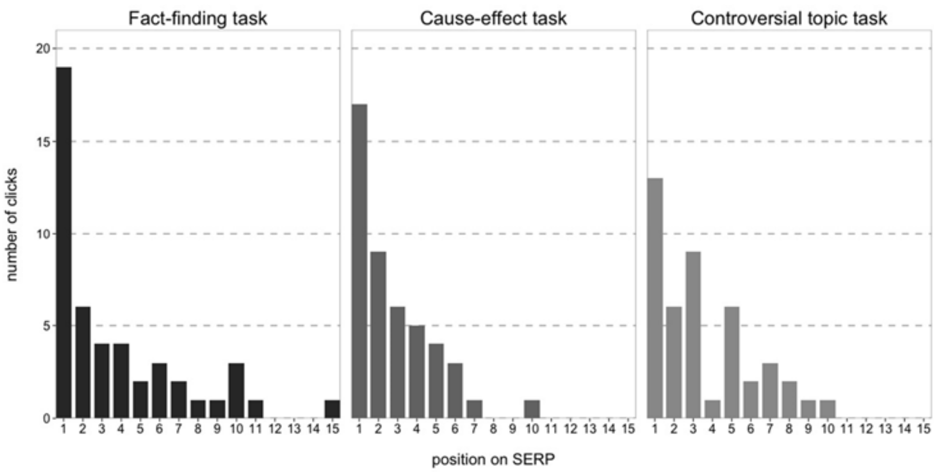

Figure 2: Position of selected webpages on the SERP

Table 4

Distribution search result selection: skewness & kurtosis		
	Skewness	Kurtosis
Fact-finding task	2.465	5.582
Cause-effect task	1.688	2.161
Controversial topic task	1.294	0.520

When looking at the *number of viewed search results*, participants considered more search results in the cause-effect task in comparison to the controversial topic task ($\chi^2(2) = 9.46$, $p < 0.01$). The other two post hoc comparisons showed no significant difference. When looking at the number of *viewed search results per query*, there is no significant difference between the tasks ($F(2, 28) = 0.97$, $p > 0.05$). Another valuable perspective is whether the *search result is viewed based on the position* on the SERP. As known from earlier research, the position of the search result matters considerably. In Figure 3 the percentage of participants looking at search results at a certain position is visualized. A score of 100% for the first search result means that the participants always look at that search result. Consequently, a lower scores means that not all participants look at search results at that specific position (e.g., 25% participants looked at least once at the last search result of the SERP in the fact-finding task). As can be seen, the search results taken into consideration decrease substantially for search results that are positioned lower on the SERP. There is a significant difference between the controversial topic task on the one hand and the other two tasks on the other hand ($\chi^2(2) = 9.46$, $p < 0.01$). The number of views declines considerably faster in the controversial topic task for search results lower on the SERP.

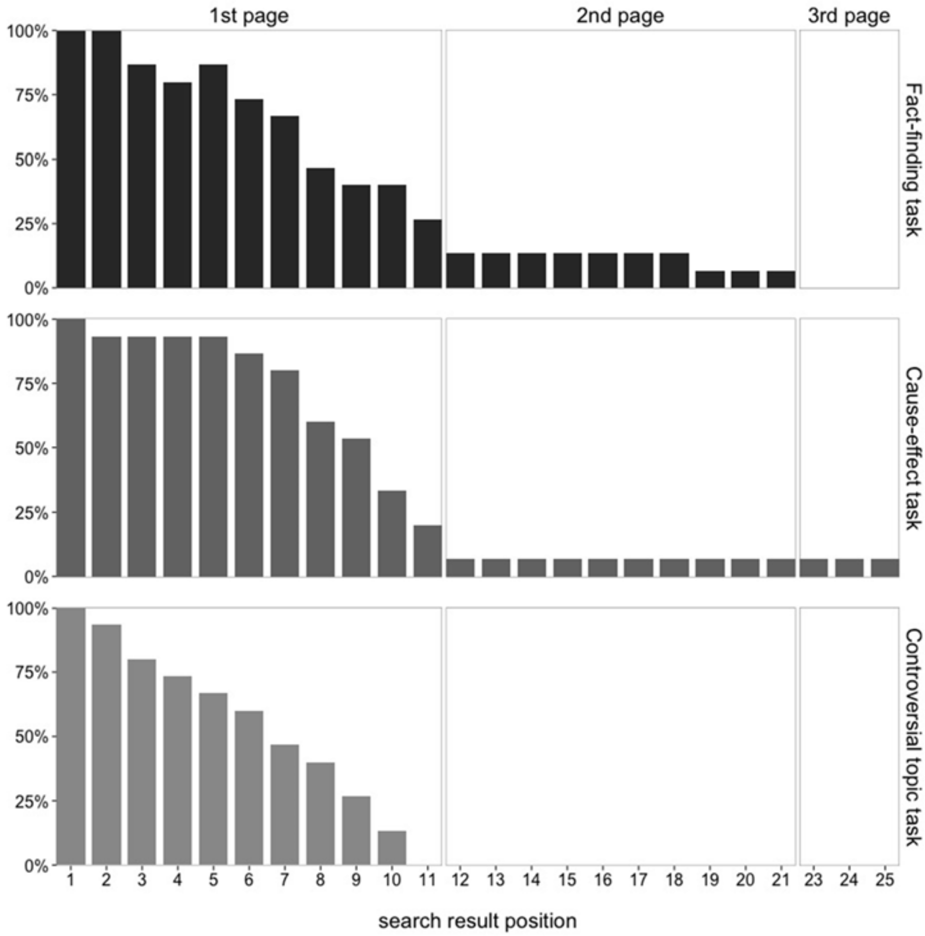

Figure 3: Considered search results by position on the SERP

The verbalized evaluative thoughts were transcribed into five categories: 1) vague evaluation, 2) the kind of webpage, 3) the web address, 4) the rank in the list of search results, or 5) reputation. There was no significant difference in the total number of evaluative thoughts between the tasks ($\chi^2(2) = 1.33$, $p > 0.05$). There was also no significant difference in type of evaluative thought between the tasks ($\chi^2(2) = 1.71$, $p > 0.05$). It seems that the participants in this study hardly evaluated the search results explicitly. Figure 4 visualizes the number of evaluative thoughts by type and per task.

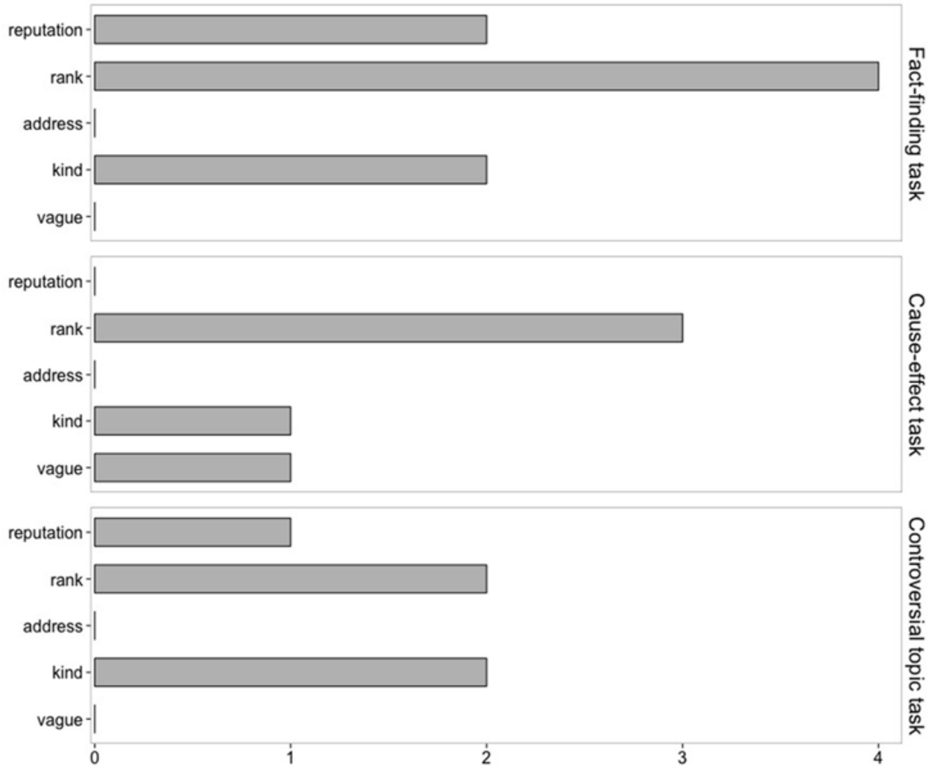

Figure 4: Number of evaluative thoughts per task

Differences in task performance

To answer the question "Does task complexity have an effect on task performance?" the task performance was analyzed by looking at the quality of the answers, time on task, and time to formulate an answer. The *quality of the answer* of the controversial topic task was significantly lower compared to the other two tasks ($\chi^2(2) = 19.70$, $p < 0.01$). The *time on task* was shorter for the fact-finding task as compared to the other two tasks ($F(2, 28) = 10.31$, $p < 0.01$, $\eta^2 = 0.248$). There was no significant difference between the cause-effect task and the conflicting topic task. The *time to formulate an answer* was significantly shorter in the fact-finding task than in the other two tasks ($F(2, 28) = 6.39$, $p < 0.01$, $\eta^2 = 0.196$). There was no significant difference between the cause-effect task and the conflicting topic task.

DISCUSSION

The present study investigated the effect of different task complexities on search behavior and task performance. The concept of search behavior was bound to the aspects of search query formulation and the evaluation of the search results on the search engine results pages. Three search tasks of differing complexity levels were studied and the used complexity levels were determined with the Mosenthal's framework of prose task characteristics (Mosenthal, 1998). The most obvious finding is that an increase in the complexity of the task influences both the quality of the outcome as well as the search behavior itself.

The results concerning the formulation of search queries revealed some discrepancies with previous research. The cause-effect task resulted in behavior that confirms previous research (Aula et al., 2010; Barsky & Bar-Ilan, 2012; Borlund & Dreier, 2014; Kelly et al, 2015; Liu et al., 2010; Singer et al., 2012a; Toms et al., 2008) which found that more time was needed to formulate the first query, more search queries were made, and more keywords per query are needed in comparison to the less complex fact-finding task. However, although the controversial topic task can be considered the most complex of the three tasks, in contrast with earlier findings, participants formulated their first query faster and used fewer queries and keywords per query as compared to the cause-effect task. Furthermore, in the controversial topic task the highest proportion of the used search keywords were derived from the task description, whereas in the cause-effect task this proportion was the lowest. The second part of this study concentrated on the evaluation of the search results on SERPs. The time spent on the SERPs and the time to make a first selection were the highest in the cause-effect task, which is in agreement with the results obtained in earlier studies (Singer et al., 2012a; Toms et al., 2008). However, both measurements were the lowest for the controversial topic task. These results indicates that the participants spent relatively less effort in evaluating the search results in the controversial topic task while addressing more attention to the actual selected webpages. The viewing and result selection patterns were in line with what was reported by previous studies (Balatsoukas & Ruthven, 2012; Höchstötter & Lewandowski, 2009; Pan et al, 2007; Savenkov, Braslavski & Lebedev, 2011): the top search results were viewed and selected more often than the ones at the bottom of the SERPs. Additionally, the participants verbalized especially the formulation of the search query and did not verbalize thoughts on the evaluation of search results very often. A similar inconsistency with regard to the controversial topic task occurs in the evaluation of the search results. Compared to the cause-effect task, the participants spent less time on the SERPs and needed less time to make a first selection in the controversial topic task.

The third question in this study sought to determine the effect of task complexity on task performance. Considering the quality of the answer, participants performed the worst on the most difficult task, i.e., the controversial topic task. This corroborates with earlier research (e.g., Byström & Järvelin, 1995). Although they performed significantly better on the cause-effect task compared to the controversial topic task, time on task and time to formulate an answer did not differ significantly between those two tasks but was longer for both tasks when compared to the fact-finding task. With regard to time on task these findings are in line with earlier research, which also showed that more complex tasks required more time (Borlund & Dreier, 2014; Brennan et al., 2014; Liu et al., 2010). Participants needed more time to formulate an answer for the cause-effect task and the controversial topic task, which might indicate the experienced difficulty in those two tasks.

Taking into account that the controversial topic task was the most complex, these results are partly contradictory to the results presented by other studies. This can be explained in several ways. First, it might be that the participants were rushing to the end of the experiment because the controversial topic task was the last one they had to execute. However, this explanation is challenged by the fact that participants spent almost the same amount of time on the task as with the cause-effect task, needed approximately the same amount of time to formulate an answer, verbalized even more thoughts when evaluating the selected webpages, and selected more webpages than in the other tasks (although not a significant difference). If participants were rushing to the end, they would have shown other behavior on these measures.

A second explanation might be the mediating effect of prior attitudes with task execution. Although we did not look into this, one might argue that already existing beliefs had an effect. As other research has shown, prior attitudes affect the information problem solving process in several ways (e.g. Brannon, Tagler & Eagly, 2007; Fisher & Greitemeyer, 2010; Van Strien, Brand-Gruwel & Boshuizen, 2014; White, 2014). Of importance to this study is the effect on keyword choice and search result selection. People tend to search for information that is consistent with their prior beliefs (Fisher & Greitemeyer, 2010) and hence formulate search queries that are strongly biased to finding supporting information (White, 2014). Furthermore, people with strong beliefs are more likely to choose information that is in line with their beliefs (Brannon, Tagler & Eagly, 2007; Van Strien, Brand-Gruwel & Boshuizen, 2014). Such prior beliefs could have prevented participants from searching critically and only looking for sources that justified their beliefs. This claim is supported by the data on the viewed search results related to the position of the search result. The decline for the lower positioned search results is much more prominent than in the other two tasks.

A third explanation is that the controversial topic task was too complex. In our assessment of the tasks with the Mosenthal framework, the controversial topic task received the highest score. What might have played a role in this experiment is the expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga, 2007). This effect explains the interaction between giving support in a task and the performance on that task. An important conclusion (Kalyuga et al., 2003; Kalyuga, 2007) is that the level of support has to be tailored to the level of prior knowledge. For search tasks this means that when students have more prior knowledge, less support should be given. In the current study no support was given. For the cause-effect task it is quite plausible that the participating students had at least some prior knowledge from their education in the physics domain as refraction indexes are addressed quite early in the curriculum, which is necessary knowledge for understanding the phenomenon of mirroring roads. This is not the case for radiation since this subject is usually not addressed at the beginning of the curriculum. In such cases the searcher's ability to determine the information goal and to evaluate the results is hampered (Bell & Ruthven, 2004). According to Ingwersen (2000) such a task is classified as a muddled topical information need in which users search for information outside familiar topics or domains. As Borlund and Dreier (2014) reported, such a muddled information need is, among others, characterized by a high number of visited web pages and a low number of search terms per query. Our study revealed similar results.

Probably both explanations play a role. There is a lot of controversy and strong opinions on the topic of mobile phone radiation. When people have beliefs that support the claim of negative health effect due to mobile phone radiation, they will unintentionally search in such a way that supporting information will come out on top of the resulting SERPs of their search queries (White, 2014). Furthermore, a lack of prior knowledge will hamper the ability to generate proper keywords for a search query, which is confirmed by the fact that the participants hardly used new keywords compared to the cause-effect task. This is supported by the research of White, Dumais and Teevan (2009), who showed that development in domain expertise results in an increase of query vocabulary. Moreover, participants spent more time on the selected webpages which might imply that they needed more time to analyze whether that page was valuable or not as a result of their lack of prior knowledge.

LIMITATIONS OF THIS STUDY

A limitation of this study is that we did not rotate the order of the tasks. Several findings indicate that this did not severely affect the outcomes of this study. Time on task for the third task was approximately the same as for the second task. Time used to formulate an answer was not shorter than in the second task. Participants spent relatively more time on the selected webpages and relatively less time on the SERPs. This is an indication that they had more difficulties to comprehend the contents of these webpages. Although not significantly different, the number of selected webpages was the highest for the last task (the controversial topic task). These findings indicate that participants were similarly engaged in all three tasks and no fatigue effect occurred.

Another limitation of this study is the relatively small number of participants. This decreases the statistical power for the measurements where the number of observations was equal to the number of participants (i.e., number of search queries, time to formulate the first search query, time to first selected search result, number of selected search results, quality of the answers, time on task, and time to formulate an answer). For the other measurements the number of observations is much higher as each participant visited several SERPs. To increase generalizability of this first set of results, replication with a larger number of participants would be desirable.

CONCLUDING REMARKS

In this study we investigated how differences in search task complexity influence different aspects of search behavior. We can conclude that increases in task complexity leads to increases in interaction with the search engine (including the SERPs) and poorer task performance. However, when a task becomes too complex, interaction drops and task performance drops, too. These findings have several implications. The most important implication of these findings for education is that search tasks should be designed such that the complexity of the task and the accompanying support is tailored to the abilities of the learners (Van Merriënboer & Kirschner, 2012).

With regard to future research, it is valuable to take both prior attitudes as well as experienced task difficulty (subjective task complexity) into account (cf. Gwidzka & Spence, 2006). As most research is conducted with the Google search engine, it would be valuable to investigate the uses of a greater diversity of search tools as brand name can have an influence on the preference of search results (Bailey, Thomas & Hawking, 2007). Another possible route to explore, is the use of extended webpage summaries in the SERPs as they enable user to make better judgments about web page relevancy (White et al., 2003). Especially for novices, support with such summaries might help them to develop a better sense about which search results are relevant for their information need. This could possibly be combined with methods like the School Assignment Satisfaction Index (SASI) as introduced by Bailey et al. (2010), which also help make better judgments about search result relevance.

The main conclusion from this study is that an increase in the complexity of the task influences both the task performance as well as the search behavior itself.

Chapter 3

Learning and navigating in hypertext: Navigational support by hierarchical menu or tag cloud?

This chapter is based on:
Walhout, J., Brand-Gruwel, S., Jarodzka, H., Van Dijk, M., De Groot, R. & Kirschner, P. A. (2015). Learning and Navigating in Hypertext: Navigational Support by Hierarchical Menu or Tag Cloud? *Computers in Human Behavior, 46,* 218-227. doi: 10.1016/j.chb.2015.01.025

Abstract

As hypertext learning environments (HLE) are widely used in education, it is important to study and know the effects and consequences of its use. HLEs are non-linear which means that students have to develop ways of navigating through them. Thus, developing interfaces that facilitate and even guide navigation is important for learning. Research showed that successful learning in HLEs depends on both learner characteristics and HLE features. This study investigated an HLE navigation feature (navigational support with either a tag cloud or conventional hierarchical menu), task complexity (fact-finding vs. information-gathering task), and a user characteristic (gender). Results show that neither navigational support nor gender is associated with differences in task performance. However, there are differences in information processing. Participants using tag clouds looked longer at the navigational support and shorter at the overview pages. Combined with fewer revisits of webpages in the tag cloud condition, this indicates a more focused selection of pages. The deeper processing of information needed for the information-gathering task was reflected in fewer visits to, but longer viewing times of pages. As no differences in task performance were found, tag clouds seem to be as effective for performance as more traditional navigation structures for navigational support.

In today's education, hypertext learning environments (HLEs) are widely used to enrich traditional education. Aleven, Stahl, Schworm, Fischer and Wallace (2003) describe a HLE as an on-demand help containing context specific hints, hyperlinked background material, textbooks, and glossaries. Such HLEs are often used for inquiry-based learning, discovery learning, web quests, and so forth; that is in environments where the instruction does not necessarily follow a set order or structure. Due to its inherent nature, the content of these HLEs is presented in *non-linear ways* making it more difficult to navigate the contents than traditional linear text (Scheiter & Gerjets, 2007). A further complication is that HLEs make use of two different kinds of interfaces: the browser interface and the website (HLE) interface (Juvina & Van Oostendorp; 2006). While the browser interface / brand of browser hardly changes and the differences between browsers is minimal, the variation in website interfaces is enormous and often unique for each website. Thus, students have to develop *new ways of navigating* for each new HLE they encounter. For instructional designers, it is therefore important to design website interfaces for HLEs with ease of use in mind, and in such a way that navigation and learning are facilitated. In this study navigation in HLEs with tag clouds is compared with navigation with hierarchical menus. A tag cloud is a visual representation of keywords in the form of a cloud. While the hierarchical menu is standard, the novel tag cloud menu is upcoming. However, little is known about its use, yet. Both Anfinnsen, Ghinea and De Cesare (2011) and Voit, Andrews and Slany (2009) indicate the usefulness of tag clouds in browsing, but stress the need for more research. This study fills this research gap by comparing hierarchical menus with tag clouds. It can be expected that design choices have an influence on navigation behavior. As these influences are not known yet, the most commonly used design choices for tag clouds will be studied.

HLEs can be characterized as a database of pieces of information (often called 'nodes') where relations between the nodes are represented by explicit links (Conklin, 1987; Kim & Hirtle, 1995). In contrast to printed text where pieces of information are organized and made available in a fixed consecutive order, HLE users can jump from one node to another without following a predefined path. One often given advantage of the latter is that users are able to determine which links to follow and in which order (Conklin, 1987) and thus can adapt the amount and type of information they consume to their information needs. As the information nodes are modular, links to each node can be created from several different other nodes. This might increase the efficiency of the learning process (Dee-Lucas & Larkin, 1999) and learner motivation (Mobrand & Spyridakis, 2007). However, a negative consequence of this 'freedom' is that as a result of the less clear structure learners might get lost and consequently show less coherent reading order of text nodes and eventually create a 'fragile' network of knowledge offering "frail and casual webs of information that lead to the cultivation of similarly flimsy mental networks (the 'Butterfly Defect')" (Salomon & Almog, 1998, p. 222). Likewise, users need to use additional cognitive processing capacity above the simple processing of the information because they have to decide where to go next and why. As more mental effort is needed to navigate through HLEs, the risk of overload of working memory will be increased as well. As Madrid, Van Oostendorp, and Puerta Melguizo (2009) concluded, a more coherent reading order of text nodes is correlated to reduced cognitive load. This indicates that design choices have an influence on cognitive load as well.

Previous research showed that the learning success from HLEs depends on learner characteristics (e.g., working memory capacity, expertise in a domain, ability to impose structure on information, gender) in conjunction with the particular features of different types of HLEs (e.g., hyperlink structure, number of nodes), the complexity and type of task, and the type of navigational support, such as hierarchical or tag cloud (DeStefano & LeFevre, 2007; Puerta Melguizo, Vidya & Van Oostendorp, 2012). This study focuses on the type of navigational support and the task complexity, taking into account gender differences as an aspect of user characteristics. As little is known about the influence of design choices of tag clouds on navigation behavior, the most commonly used design choices for tag clouds will be studied.

Navigational support: Hierarchical or tag cloud

Guiding learners while navigating through HLEs might reduce the cognitive burden of continuous decision-making, help them build a coherent mental structure of the information, and help them to use more coherent and efficient navigation behaviors thus facilitating learning. Not using such support might, in contrast, influence learning outcomes negatively (Minetou, Chen & Liu, 2008). However, it is not clear if this is actually the case and if so, which type of navigational support should be used and how this support best can be designed.

Previous research on HLEs compared different types of graphical overviews (e.g., Amadieu et al, 2009; Bezdan, Kester & Kirschner, 2013) or different types of hierarchical menus (e.g., Leuthold, Schmutz, Bargas-Avila, Tuch & Opwis, 2011; Puerta Melguizo et al., 2012). Generally, one can conclude that a hierarchical organizer is slightly better for navigation compared to a network organizer (Amadieu et al., 2009), using dynamic menus hinders task performance (Bezdan et al., 2013), and it is not wise to severely restrict learner navigation paths (Bezdan et al., 2013). Furthermore, vertical menus are preferred over dynamic menus (Leuthold et al., 2009). These studies show clearly that design choices have an influence on user performance. It is therefore important to study the effect of these design choices.

In a hierarchical system, information is organized in a folder structure where folders can contain subfolders and so further, which results in a tree-like structure. Recently, tag clouds have come into use as a new tool for navigation. Examples of this tagging approach are Delicious®, Pinboard® (favorite websites), Youtube® (video), Flickr® (photo's), Connotea® (science) or Last.fm® (music). Few such studies exist (Trattner, Lin, Parra, et al,

2012) investigating how to construct and visualize tag clouds. The same is true for comparing the use of tag clouds and more traditional hierarchical menus (Civan, Jones, Klasnja & Bruce, 2008; Voit, Andrews & Slany, 2012).

Most participants reported that using tags is a better and more flexible way to search for information (Civan et al., 2008). The greater flexibility and the descriptive characteristics of tag clouds were appreciated. Research about filing and re-finding behavior with hierarchies and tagging systems did not find significant differences between tags or folder hierarchies (Voit et al, 2012). However, fast performers required less time and fewer mouse clicks when using tags while slow performers benefitted more from hierarchical folders. Feedback from the participants indicated that the use of tags improved the subjective user experience, while requiring an equivalent amount of time. However, these studies investigated only user behavior, but not the cognitive and perceptual processing leading to this behavior. Whether the use of tags instead of hierarchical menus has an effect on learning outcomes and its underlying processes is unknown, but the findings of Civan et al. (2008) and Voit et al (2012) indicate that the use of tags can be a viable alternative for navigational support in HLEs.

Task complexity

Experiments are frequently designed in such a way that the requested information can be found in one specific page of a certain website, although, in reality information is often scattered over various pages and websites (Puerta Melguizo et al., 2012). Hence, there is a distinction between ‘fact-finding tasks’, in which the information can be found in a specific place, and ‘information-gathering tasks’ where participants have to gather and combine information from different sources in order to find an answer. Information-gathering tasks are more difficult because collecting and integrating information from different sources requires that pieces of information be remembered while continuing the search task (Rouet, 2003). Consequently, cognitive load is higher in information-gathering tasks. As hierarchical menus present a more coherent reading order, they might be better suited for such complex tasks.

Individual characteristics: gender differences

Research found that boys are more active in browsing than girls (Large, Bahesti & Rahman, 2002; Roy & Chi, 2003): they looked at more pages, selected more hyperlinks and saved more information, while they spent less time on viewing pages. It can be expected that learning outcomes are affected negatively as spending less time viewing pages might hamper comprehension. Moreover, female students have been found to have more difficulties orienting and navigating (Ford, Miller & Moss, 2001). Female students also felt less in control and experienced feelings of getting lost more often. On the other hand, Hupfer and Detler (2006) found no evidence of gender differences in navigation behavior.

Based on these gender differences we can only speculate which type of navigational support suits females better than males or vice versa. However, spatial ability research can give us a direction.

Research found that when people talk about navigating through hypertext, they mostly use terms that are also used for spatial navigation (Kim & Hirtle, 1995; Maglio & Matlock, 1998; Hochmair & Luttich, 2006). When verbalizing our thoughts about navigating on the Internet, phrases as "I *went to* this webpage", "I found that *at* Wikipedia" or "The *address* of this website" are often used. We also see this in the browser interface, which uses buttons labeled as "*home*", "*back*" or "*forward*". As navigation tasks require participants to move throughout cyberspace, spatial ability might be an important factor in determining success and/or difficulty of HLE navigation. In line with this argumentation, some research has shown a link between spatial ability and HLE navigation (Campagnoni & Ehrlich, 1989; Juvina & Van Oostendorp, 2006; Stanney & Salvendy, 1995). These studies showed that high spatial ability is connected to better and faster task performance on information retrieval tasks. This more efficient performance is characterized by visiting less non-relevant pages, using the back-button less often and lower experiences of feeling lost (Ahmed & Blustein, 2005). For gender differences in learning, memory and spatial ability see Andreano and Cahill (2009).

HYPOTHESIS

The experiment presented in this article focuses on the effects of two types of navigational support, information tasks with differing complexity levels, and learner gender on navigation behavior through HLEs, visual processing of the navigation menu (as measured by eye tracking), and task performance.

With respect to *navigational support,* this study compares tag clouds as a navigational support system with the more commonly used hierarchical menus. Based on the review of the literature in the previous section, it can be expected that participants using tag clouds will visit fewer pages because they will be able to locate the desired webpages more accurately (H1a). Consequently, based on the assumption of visiting fewer pages, it is likely that the use of a tag cloud will result in a faster task performance compared to the use of a hierarchical menu (H1b). Moreover, as a result of the higher information density of tag clouds, it is plausible that the participants in the tagging condition will need more viewing time for the navigation structure than participants using a hierarchical menu (H2a). This may neutralize the effect of visiting fewer pages (H2b). Furthermore, as tag clouds have a higher information density and are more cluttered, it could involve more time to take decisions (H2c). On the one hand, tag clouds give a more detailed description of the contents of the HLE which might help participants better locate the desired information. We therefore expect that participants in the tagging condition will carry out the tasks (H3a) in a better way. On the other hand, as Madrid, Van Oostendorp, and Puerta Melguizo (2009) concluded, a more coherent reading order of text nodes is correlated to reduced cognitive load. Thus, participants in the hierarchical condition will give better answers (H3b).

The *complexity of a task* can have an effect as well. A *gathering task* requires more mental effort and deeper processing of the texts because learners need to synthesize information from several pages into an answer. Whereas for the *fact-finding task*, learners only need to find the correct page with the desired information. Therefore, we expect to see several differences between those two tasks. In the fact-finding task most time is probably spent on deciding which page to go to. We therefore expect to see more 'trial and error' behavior, which is characterized by relatively short page visits to more pages (H4). As locating the desired information is such an important part of the fact-finding task, we expect that more attention is given to the navigational support and in the gathering task the texts will receive more attention (H5).

Finally, as discussed in the previous section, from the gender perspective it can be expected that boys will visit more pages, view them for shorter periods of time than girls and will find the desired answers more quickly (H6). Hypotheses 4, 5, and 6 will be investigated in relation to the two types of organizers.

METHOD

Participants and Design

To study the influence of navigational support and gender in tasks of different complexity levels on navigation behavior and task performance in hypertext learning environments, a 2 x 2 factorial design with a within subject measurement was used with the factors *type of navigational support* (hierarchical menu vs. tag cloud) and *gender* (male vs. female) as between subject factors and *task complexity* (fact-finding task vs. gathering task) as a within-subjects factor. The dependent variables were task performance and navigation behavior.

A total of 60 tenth grade students (M_{age} = 15.63 years, SD = 0.688) in the highest level of secondary education of in the south part of The Netherlands participated. 24 boys (M_{age} = 15.75, SD = 0.737) and 36 girls (M_{age} = 15.56, SD = 0.652) were involved. Participants were equally divided across the two conditions of navigational support (i.e., tag cloud vs. hierarchical menu). Participants had experience in using their school's HLE, but had no experience in using other HLEs. Participants had normal or corrected to normal vision.

Material and apparatus

Task. Two tasks on the topic of 'obesity' were constructed. This is a topic that students are familiar with, but that is not part of the regular school curriculum. The first task was a *fact-finding task* in which the students had to find the percentage of obese people in the Netherlands, which is 14 percent. The second task was an *information-gathering task* where students had to give three reasons why obesity is a bigger problem in lower social-economic classes of society than in higher ones, which are: (1) healthy foods are more expensive, (2) higher educated people have more knowledge about a healthy diet and (3) minorities (which are over-represented in the lower social-economic classes) consume more fastfood. Students located and stored relevant information while navigating through pages in either the hierarchical or the tagging hypertext environment. The students received three minutes to execute the simple fact-finding task and seven for the complex information-gathering task.

This study focuses on differences between two levels of task complexity. In order to describe the difference in complexity between the tasks, the prose task characteristics of Mosenthal (1998) were used. Task complexity can be judged along three dimensions: type of information requested, type of match, and plausibility of distractors. The type of information requested relates to how concrete or how abstract the information is that must be identified. Questions about identifying an amount or percentage are considered to be easier to solve than questions requesting causes. Type of match relates to the similarity between the information in the question and the information in the text and the processes needed to connect those two. The cause-effect relationship requested in the *information-gathering task* is considered more difficult than locating a feature or characteristic as in the *fact-finding task*. The plausibility of distractors describes the degree in which other pieces of information have similar characteristics in relation to the requested information. The more distractors and the closer they are located to the requested information in the text, the more difficult a task is.

Based on these three task characteristics, Mosenthal (1998) describes five levels of task difficulty. The *fact-finding task* can be assessed as a task of difficulty level one, whereas the *information-gathering task* is a task of difficulty level three.

Hypertext environments. Two computer-based e-learning environments were set up on the topic of obesity, one for each condition. Both HLEs were built with the open source software Wordpress®. The environments for both conditions were almost identical; the only difference was the navigational structure. As the research presented in paragraph 2.1 shows, design choices affect the effectiveness and efficiency of users in their navigation behavior. As the influences of design choices of tag clouds are not known yet, it is appropriate to compare the most commonly used design choice for tag clouds with hierarchical menus. This results in a more ecological valid setting for the design of the tag clouds as well.

The navigational support was positioned at the left side of the screen, either as an hierarchical menu (Figure 1) or a tag cloud (Figure 2). The hierarchical menu showed the categories in a hierarchy with subcategories one level deep. In this study we decided for a static and vertical hierarchical menu. The tag cloud presented a collection of tags (i.e., labels). Relationships between the tags were not shown.

The environment consisted of two types of web pages: overview pages and text pages. A total of 32 text pages were created. These text pages were categorized in a hierarchical structure and a tag cloud by one of the researchers. The other researchers then reviewed this categorization. Clicking on a menu item in the hierarchical menu or on a tag in the tag cloud provided the participant with an overview page in that specific category or labeled with that specific tag with a linear list of the corresponding page title for the selected category or tag. The participants could access the text pages by clicking on the titles at the overview pages. Each environment consisted of 32 text pages. The text pages in the environment were copied from existing Internet resources. A text page consisted of the following elements: title, information about the source, plain text without hyperlinks and sometimes a table and a button with which the page could be saved if appropriate for answering the question. Figure 1 presents an example.

Figure 1: HLE with hierarchical navigation

Figure 2: HLE with tag cloud navigation

Eye tracking equipment. Eye movements and logging data were recorded with a remote Tobii 1750 eye tracker with temporal resolution of 50 Hz (2003), which is integrated with a PC screen, and is operated with Studio software (see www.tobii.com) from the stimulus PC. The screen capture recording mode was used, so not only the eye movements, but the entire task performance process (including possible mouse and keyboard operations) was captured.

Measurements

Task performance was measured as the correctness of the answers to the questions in the tasks. Task performance on the fact-finding task was coded as correct or incorrect. For the information-gathering task, the number of correct reasons why obesity is a bigger problem in lower socio-economic classes of society was counted. The maximum number of reasons was three. *Time on task* was used as a control variable.

To gain insight in navigation behavior, logging and eye tracking data were recorded. While participants carried out the tasks, several actions (page id, start and end timestamp for each visited page) within the HLE were tracked, recorded and stored in a log file for each participant. Based on these log files, the following variables were calculated.

Time to first click. Time it took participants before they clicked on an item in the navigational support for the first time to visit an overview page. This included viewing the navigational support, making a choice, and clicking on a navigation item with the mouse.

Visited pages. Total number of visited pages was logged. From these logs, the number of *uniquely visited* pages and the number of *revisited* pages were calculated.

Time spent on overview pages and text pages. How long a participant stayed at each page. From these data, time spent on overview pages and time spent on text pages was calculated.

Eye tracking parameters were calculated for all pages in the HLE. All parameters were assigned to certain elements of these pages, so-called *areas of interest* (AOIs). A distinction was made between an AOI for the navigational support and an AOI for the content part of the pages. On the overview pages the list AOI covered the list of corresponding pages for the clicked tag or hierarchical category. On the text pages the content AOI covered the title and text of the page. All analyses were performed with Tobii Studio software version 2.2.4 (2007). Before further analyses were executed, the raw data was filtered with the Tobii ClearView fixation filter. Based on visual inspection of the data, a fixation definition of 30 pixels and 100 milliseconds was chosen (cf. Hegarty & Just, 1993; Loftus, 1981).

To analyze these AOIs, the total fixation duration was calculated. The fixation duration measures the sum of the durations for all fixations within an AOI or an AOI Group. If during the recording the participant returns to the same media element then the new fixations on the AOI will be included in the calculations of the metric (Tobii, 2007).

Procedure

Prior to the experiment, participants received an introduction to the study. Also, all potential participants received a letter for their parents in which the study was explained. The experiment was conducted in individual sessions. Each session started with a detailed explanation of the procedure. For each participant, demographic data (i.e., age, gender) were collected. Next, they were introduced to the eye tracking equipment. The eye tracker was adjusted to the individual features of each participant by calibrating the system with a 9-point calibration. The maximum duration of the sessions was 50 minutes. Participants were randomly assigned to the two different conditions.

After a short instruction, participants first had to perform the fact-finding task and then the information-gathering task. Each task started with a question after which the participants searched in the hypertext environment at their own pace and in their own order. The maximum duration for the fact-finding task was three minutes and for the information-gathering task seven. When the participant thought they could answer the question, they could stop the recording themselves by pressing the spacebar. When the maximum duration was reached, the eye tracking software stopped the session. The whole procedure required maximally thirty minutes (including instruction for each task, calibration of the eye tracker for each task, make-up removal, and a short break between the two tasks). For each participant, both tasks were recorded in one session.

Results

The results of this study are presented in three parts: task performance, logged activities and eye movements. All relevant means and standard deviations are summarized in Table 1 (fact-finding task) and Table 2 (information gathering task). Two-way factorial repeated measures ANOVAs were used to analyze the data. As part of the eye tracking data and the logging data were not normally distributed, thus, non-parametric statistical tests (Wilcoxon rank sum test and Kruskal-Wallis test for between subject factors; Wilcoxon signed rank test for the within-subjects factor) were used to analyze these data when transformation did also not result in normally distributed data.

Table 1 Means and Standard Deviations for Eye Tracking and Logging Data for Task 1

	Hierarchical navigation support			Tag cloud navigation support		
	Male	Female	Total	Male	Female	Total
N	12	18	30	12	18	30
Control variable						
Time on task	137.33 (56.31)	131.67 (47.78)	133.94 (50.49)	142.46 (51.10)	122.50 (51.49)	130.48 (51.41)
Logging						
Time to first click	7.78 (2.31)	7.18 (2.19)	7.42 (2.22)	7.34 (2.35)	7.19 (1.84)	7.25 (2.02)
Time on overview pages	77.74 (52.64)	63.40 (38.96)	69.14 (44.63)	65.15 (34.26)	48.29 (34.19)	55.03 (34.66)
Time on text pages	43.06 (29.96)	51.35 (30.18)	48.04 (29.85)	55.05 (24.86)	53.27 (24.61)	53.98 (24.30)
Visited pages	12.17 (6.09)	11.78 (5.06)	11.93 (5.39)	9.00 (3.98)	9.17 (6.81)	9.10 (5.76)
Revisited pages	3.75 (3.84)	3.06 (2.31)	3.33 (2.97)	0.83 (1.27)	1.28 (2.40)	1.10 (2.01)
Eye tracking						
Fixation duration navigation	33.35 (16.72)	31.99 (17.34)	32.53 (16.82)	54.98 (28.21)	44.70 (26.48)	48.81 (27.19)
Fixation duration overview pages	55.32 (37.73)	45.28 (26.08)	49.30 (31.04)	38.36 (19.69)	28.04 (16.80)	32.17 (18.41)
Fixation duration text pages	34.40 (24.05)	37.68 (23.66)	36.37 (23.46)	34.57 (13.56)	37.56 (19.13)	36.36 (16.93)

Table 2 Means and Standard Deviations for Eye Tracking and Logging Data for Task 2

	Hierarchical navigation support			Tag cloud navigation support		
	Male	Female	Total	Male	Female	Total
N	12	18	30	12	18	30
Control variable						
Time on task	298.28 (100.39)	302.71 (112.21)	300.94 (105.87)	297.99 (99.40)	312.06 (112.31)	306.43 (105.79)
Logging						
Time to first click	8.42 (1.94)	7.69 (2.06)	7.98 (2.01)	7.79 (1.16)	8.35 (3.06)	8.12 (2.46)
Time on overview pages	76.39 (38.58)	72.26 (43.51)	73.91 (40.97)	67.40 (31.91)	86.82 (65.84)	79.05 (54.96)
Time on text pages	206.93 (71.62)	216.98 (79.97)	212.96 (80.92)	212.20 (78.92)	204.18 (84.34)	207.39 (80.92)
Visited pages	19.50 (7.88)	18.67 (9.56)	19.00 (8.79)	17.17 (4.53)	20.22 (11.33)	19.00 (9.24)
Revisited pages	7.42 (3.42)	6.89 (5.93)	7.10 (5.01)	3.25 (2.38)	4.61 (3.42)	4.07 (3.07)
Eye tracking						
Fixation duration navigation	37.84 (19.52)	41.20 (19.64)	39.85 (19.32)	66.24 (29.02)	76.40 (48.65)	72.33 (41.62)
Fixation duration overview pages	67.01 (35.09)	57.47 (32.98)	61.29 (33.57)	47.06 (20.80)	52.80 (36.18)	50.50 (30.65)
Fixation duration text pages	181.92 (63.73)	185.02 (65.96)	183.78 (63.98)	168.05 (70.21)	157.49 (63.43)	161.72 (65.24)

Task performance

Time on task was used as a control variable in this study. For the fact-finding task Wilcoxon rank sum tests revealed no significant differences across genders ($W = 352.5$, $p = .233$), or across type of organizer ($W = 427$, $p = .739$). A Kruskal-Wallis test with four groups (male-hierarchical, male-tagcloud, female-hierarchical and female-tagcloud) did also not show a significant difference between the groups ($H(3) = 2.706$, $p > .05$) on time on the fact-finding task. For the information-gathering task Wilcoxon rank sum tests revealed no significant differences across genders ($W = 443$, $p = .874$), or across type of organizer ($W = 429$, $p = .762$). A Kruskal-Wallis test with four groups did also not show a significant difference between the groups ($H(3) = 0.124$, $p > .05$) regarding time on the information-gathering task.

Task performance is related to hypotheses H1b, H3a, H3b and H6. The numbers of correct answers are summarized in Table 3. Logistic regression analysis revealed no statistical significant relationship between task performance in the fact-finding task and type of navigational support, gender or the interaction between type of navigational support, and gender (see Table 4). For the gathering task, regression analysis revealed also no statistical significant relationship between task performance and type of navigational support, gender or the interaction between type of navigational support and gender (see Table 5).

Table 3 Task performance

	Hierarchical navigation support			Tag cloud navigation support		
	Male	Female	Total	Male	Female	Total
N	12	18	30	12	18	30
Fact-finding task						
Correct answer	5	8	13	7	11	18
Information-gathering task						
One correct reason	2	4	6	1	3	4
Two correct reasons	6	7	13	7	8	15
Three correct reasons	4	7	11	4	7	11

Table 4 Logistic regression fact-finding task: Variables in the Equation

				95% CI for odds ratio		
	B	S.E.	Sign.	Lower	Odds ratio	Upper
Constant	-.22	.47	.64	.30	.80	2.03
Gender	-.11	.75	.88	.20	.89	3.92
Navigation	.68	.68	.32	.53	1.96	7.69
Gender by Navigation	-.00	1.07	1.00	.12	1.00	8.35

Note: pseudo R^2 = .021 (Hosmer and Lemeshow), 0.28 (Cox and Snell), 0.37 (Nagelkerke).
Model χ^2 (3) = 1.72, p > .05

Table 5 Regression information-gathering task for predicting task performance

Variable	B	SE(B)	β
Navigation	0.06	0.24	.23
Gender	0.00	0.27	.00
Navigation x gender	0.03	0.38	.07
ΔR2	.05		

*$p < .05$. **$p < .01$.

Differences in behavior

Behavioral differences of the participants in this study are divided in differences in durations, differences in page visits, and differences in viewing behavior. All relevant means and standard deviations are summarized in Table 1 (fact-finding task) and Table 2 (information gathering task). Relevant test statistics are summarized in Table 6 (Analysis of Variance tests), 7 (Wilcoxon rank sum tests) and 8 (Kruskal-Wallis tests).

Table 6 Analysis of Variance (ANOVA) between navigational support, task complexity and gender

	df	F	η^2	*p*
Time to first click				
navigation support	1	0.045	0.001	> .05
task complexity	1	4.363	0.072	.041
gender	1	0.233	0.004	> .05
error	56			
Time spent on overview pages				
navigation support	1	1.367	0.024	> .05
task complexity	1	52.507	0.484	.041
gender	1	0.967	0.017	> .05
error	56			
Time spent on text pages				
navigation support	1	2.101	0.036	> .05
task complexity	1	55.705	0.499	< .001
gender	1	1.101	0.019	> .05
error	56			
Number of visited pages				
navigation support	1	4.309	0.071	.043
task complexity	1	29.873	0.348	< .001
gender	1	1.101	0.019	> .05
interaction: navigation support x task complexity	1	10.164	0.154	.040
error	56			
Fixation duration navigation				
navigation support	1	41.827	0.428	< .001
task complexity	1	67.717	0.547	< .001
gender	1	0.012	0.000	> .05
error	56			
Fixation duration overview pages				
navigation support	1	21.111	0.274	< .001
task complexity	1	8.562	0.133	.005
gender	1	0.706	0.012	> .05
error	56			

Table 7 Wilcoxon rank sum tests for navigational support and gender

	W	*r*	*p*
Number of revisited pages (fact-finding task)			
navigation support	691	-0.473	< .001
gender	471	-0.076	> .05
Number of revisited pages (information-gathering task)			
navigation support	676	-0.431	< .001
gender	385	-0.092	> .05
Fixation duration main part text pages (fact-finding task)			
navigation support	540	-0.205	> .05
gender	346	-0.146	> .05
Fixation duration main part text pages (information-gathering task)			
navigation support	651	-0.383	.003
gender	419	-0.024	> .05

Table 8 Kruskal-Wallis tests for interaction between navigational support and gender

	df	H	*p*
Number of revisited pages			
fact-finding task	3	13.380	.003
information-gathering task	3	13.108	.004
Fixation duration main part text pages			
fact-finding task	3	5.445	> .05
information-gathering task	3	9.991	.018

Task complexity had the most prominent effect on behavior, closely followed by the type of navigational support. The more complex information-gathering task resulted in more attention to the text. The tag cloud required more processing time and more effective selection of pages. Hardly any gender differences were found.

Differences in durations

Hypotheses H2c, H4 and H6 are related to differences in duration. *Time to make a decision* (hypothesis H2c) was measured by time to first click. The participants needed more time to make a decision in the information-gathering task compared to the fact-finding task ($F(1, 56) = 4.363$, $p = .041$, $\eta^2 = 0.072$). As the effect size is rather small, the effect of task complexity on the time to make a decision is not very distinct. No significant differences were found for type of navigational support, gender or an interaction between type of navigational support, gender, and task complexity.

In the fact-finding task the participants spend significantly more time on the overview pages ($F(1, 56) = 52.507$, $p = .041$, $\eta^2 = 0.484$) and less time on the text pages ($F(1, 56) = 55.705$, $p < .001$, $\eta^2 = 0.499$) when compared to the information-gathering task. Task complexity has a medium effect size on page visit duration. No significant differences were found for type of navigational support, gender or an interaction between type of navigational support, gender and task complexity.

Differences in page visits

Hypotheses H1a, H2b, H4, H5 and H6 are related to differences in the number of page visits. Regarding the *number of total page visits*, the participants visited significantly more pages in the hierarchical condition compared to the tagging condition ($F(1, 56) = 4.309$, $p = .043$, $\eta^2 = 0.071$) and significantly more pages in the fact-finding task compared to the information-gathering task ($F(1, 56) = 29.873$, $p < .001$, $\eta^2 = 0.348$). Moreover, a significant interaction between type of navigational support and task complexity was found ($F(1, 56) = 10.164$, $p = .002$, $\eta^2 = 0.154$). Pairwise comparisons with Bonferroni adjustment for this interaction showed that especially participants in the hierarchical condition visited relatively more page pages in the fact-finding task than in the information-gathering task or participants in the tagging condition.

Regarding the *number of revisited pages*, participants in the hierarchical condition did revisit significantly more pages than participants in the tagging condition for both the fact-finding task ($W = 691$, $p < .001$, $r = -0.473$) and the information-gathering task ($W = 676$, $p < .001$, $r = -0.431$). The participants revisited significantly fewer pages in the fact-finding task than in the information-gathering task ($Z = 273$, $p < .001$, $r = -0.521$). A possible interaction between type of navigational support and gender was tested with a Kruskal-Wallis test with four groups, which did show a significant difference between groups for the fact-finding task ($H(3) = 13.830$, $p = .003$) as well as the information-gathering task ($H(3) = 13.108$, $p = .004$). In the fact-finding task, female participants did revisit more pages in the hierarchical condition than in the tagging condition or male participants in the tagging condition. In the information-gathering task, male participants did revisit relatively more pages in the hierarchical condition than in the tagging condition or female participants in the tagging condition.

Differences in viewing behavior

Hypotheses H2a, H2c, H5 and H6 are related to differences in viewing behavior. The *attention given to the navigational support* was significantly higher in the tagging condition than in the hierarchical condition ($F(1, 56) = 41.827$, $p < .001$, $\eta^2 = 0.428$). Furthermore, the participants gave significantly more attention to the navigational support in the fact-finding task than in the information-gathering task ($F(1, 56) = 67.717$, $p < .001$, $\eta^2 = 0.547$). No significant differences were found for gender or an interaction between type of navigational support, gender and task complexity. Type of navigational support as well as task complexity has a medium effect on the attention given to the navigational support.

The *viewing time devoted to the main part of the overview pages* was significantly shorter in the tagging condition than in the hierarchical condition ($F(1, 56) = 21.111$,

$p < .001$, $\eta^2 = 0.274$). Regarding task complexity, the participants looked significantly shorter on the main part of the overview pages in the fact-finding task compared to the information-gathering task ($F(1, 56) = 8.562$, $p = .005$, $\eta^2 = 0.133$). No significant differences were found for gender or an interaction between type of navigational support, gender, and task complexity. Type of navigational support as well as task complexity has a medium effect on the time devoted to the main part of the overview pages.

Dedicated *viewing time for the main part of the text pages* was significantly shorter in the tagging condition than in the hierarchical condition in the information-gathering task ($W = 651$, $p = .003$, $r = -0.383$). No significant differences were found for gender or for type of navigational support and gender in the fact-finding task. Furthermore, participants did look significantly shorter on the main part of the text pages in the fact-finding task than in the information-gathering task ($Z = 307$, $p < .001$, $r = -0.577$). A possible interaction between type of navigational support and gender for the information-gathering task was tested with a Kruskal-Wallis test with four groups, which did also show a significant difference between groups ($H(3) = 9.991$, $p = .018$). The male participants in the hierarchical condition did look relatively longer on the main part of the text pages than the male participants in the tagging condition.

DISCUSSION

The main question of this study was whether it would make a difference using a hierarchical menu or tag cloud for navigation behavior and learning and how this is related to task difficulty (simple fact-finding vs. complex information-gathering) and learner characteristics (i.e., gender). Critical was whether task performance was affected or not. Based on the assumption that users of a tag cloud will visit fewer pages, we predicted that the use of a tag cloud would result in better task performance for participants with tag cloud navigation compared to hierarchical menu navigation (H1b, H3a and H3b). This expectation was not supported by the results. Although the relative frequency of correct answers was higher on the fact-finding task in the tagging condition, there were no significant differences between the two types of navigational support with either of the tasks. There were also no significant differences in task performance with regard to task difficulty or gender (H4 and H6). As there were also no differences in time on task, the results show neither negative nor positive effects of the use of a non-traditional type of navigational support (i.e., a tag cloud). One of the most intriguing findings concerning task performance is that, contrary to our expectations, users in the tagging condition did not need more time to make decisions (H2c). This could be due to the better descriptive value and better navigational flexibility of tag clouds (Civan et al., 2008) compared to a hierarchical menu.

Regarding navigation behavior, we predicted that participants who received a tag cloud would visit fewer pages compared to participants receiving a hierarchical menu

(H1a) because they would be able to locate desired webpages more accurately (e.g., Voit, et al., 2012). What was found was that participants in the hierarchical condition visited relatively more pages in total and also revisited relatively more pages than participants in the tag cloud condition. As there were no differences in time on task, it can be concluded that the use of tag clouds as navigational support leads to browsing behavior that is less active. Moreover, the smaller number of revisited pages indicates more effective localization of the desired information. As a result of the higher information density of tag clouds, we expected that the participants in this condition would need more viewing time when using the navigational support (H2a and H2b). Consequently, it might be the case that they consider what to choose more deeply. As the results showed, the participants in the tagging condition viewed the tag cloud considerably longer compared to the participants who used the hierarchical menu. This was accompanied by shorter viewing times on the main part of the overview pages. In combination with the fewer page revisits, this indicates that the use of tag clouds may lead to more focused page selection and better processing of the navigational support compared to a hierarchical menu.

With respect to task complexity, we expected more active browsing behavior for the information-gathering task compared to the fact-finding task (H4). The results, however, showed partly the opposite. Participants visited relatively more (unique) pages in the fact-finding task, but at the same time revisited more pages in the information-gathering task. This latter effect could be because an information-gathering task requires more elaborate reading of the text in order to be able to carry it out than a fact-finding task. This is supported by the result obtained that participants spent relatively more time on the text pages in the information-gathering task (H5). Moreover, the longer decision time to the first click is also an indication that the information-gathering task required more elaborate reading. Finally, the higher frequency of page visits in the fact-finding task might point to a more trial-and-error type of navigation behavior.

Although most of the results did not reveal differences, some minor differences between the genders were found. These findings were contradictory to the expectations based on earlier research (Large, Bahesti & Rahman, 2002; Roy & Chi, 2003). While both genders needed fewer revisits when given a tag cloud, males profited more from tag clouds than females. Thus, the use of tag clouds leads to less active browsing behavior, in particular for males. On the other hand, males in the tagging condition viewed the texts for shorter periods of time, which might be an indication of more active browsing behavior.

Since students are most used to receiving and using hierarchical structures in their learning, navigating with the use of a hierarchical menu is less demanding for them than navigating with less familiar navigation structures such as the tag clouds used in this study. Although they had to carry out two tasks, this was probably not enough time to get used to using tag clouds for navigation. As such tags are valued for their descriptive value and are perceived to be slightly more favorable than hierarchical structures (Civan et al., 2008), when the use of tags becomes more common, users might be able to use them more efficiently.

Finally, nearly every design decision influences behavior. It is therefore likely that our design decisions have influenced behavior. However, our design decisions ware based on common practice in constructing hierarchical menu structures and tag clouds.

DIRECTIONS FOR FUTURE RESEARCH

The most important conclusion is that tag clouds can be considered as a viable option for navigational support. As little is known about the influence of design choices of tag clouds on navigation behavior, this study is therefore a first step in studying the use of tag clouds as navigational support in HLEs. However, more research is needed concerning the different design choices of tag clouds. Studying variations in the design of the navigational support will further develop the understanding of the use of tag clouds. For example, varying the level of detail or varying the visual presentation will give more insight into the influence of specific design decisions on behavior and performance. A more thorough understanding will also result in sound recommendations for the design of HLEs.

In this study two task were used, a fact-finding task and a more complex information-gathering task. It has to be noted, though, that while the information-gathering task was more complex than the fact-finding task, neither task was really very complex. Another future research direction is to look at other, more complex tasks and maybe also other task types. An example of a more complex task is an information-comparison task. In such a task the learner not only has to find and gather information, but also has to determine relevant dimensions on which to compare the information and then actually carry out the comparison. A step further is an information-evaluation task where after the comparison decisions need to be made as to value, use for a certain purpose, et cetera. Examining navigation behavior for other, more complex, tasks is necessary to build a more coherent understanding of different types of navigational support. Moreover, future research should make use of other task topics as well.

Another aspect for future research is the size of the HLE's. In our study, the HLE was limited to 32 pages. In reality, HLE's can contain more and sometimes almost endless amounts of content (e.g., Wikipedia®). It would be interesting to see whether our conclusions are applicable in large scale HLE's. In this study we choose for a specific lay out, but other layouts might induce other effects (Lohman, Ziegler & Tetzlaff, 2009).

Furthermore, in this study we choose gender as the individual characteristic to investigate. In future studies it would be interesting to see whether other individual characteristics (e.g., working memory capacity, expertise in a domain, ability to impose structure on information) do have an influence.

CONCLUDING REMARKS

Navigating in hypertext learning environments is a complex process; one that needs to be understood much better by instructional designers and educational practitioners (e.g., teachers). The choice of a menu style and item can have far reaching implications. Choosing the wrong type for the task at hand or for a specific type of learner may, for example, slow a student down because (s)he selects a wrong menu item or hyperlink. As a result, the reading and information processing is interrupted, which in turn might hamper the learning process. The results of this study show that learning outcomes can benefit from using a novel tag cloud as navigational support as compared to more traditional navigation structures. Tag clouds can unfold their full potential, in particular for large HLEs, making large amounts of information easier accessible. Moreover, the descriptive nature of tag clouds could enhance the information processing as well.

Chapter 4

To tag or not to tag? How to support organizing and classifying bookmarks from the web?

Abstract

When searching for information on the Internet, people often find information they want to keep for later use. Because they want to re-access this information somewhere in the future, the question arises how to keep this information. Moreover, people want to keep information about different topics. Using a bookmark system is a possible way to support such needs. However, the ability to keep and organize this information depends for a large part on prior knowledge as this is needed to come up with a classification schema. A lack of prior knowledge hinders novices in their process. Giving them support might overcome such a barrier. In this study a between-subjects design was used to study the influence of content support with a classification schema and tool support with either a tagging or a hierarchical system. The results showed that despite the low familiarity with tagging systems, using a hierarchical system is not necessarily a better approach than using a tagging system to organize previously found information.

Since the rise of the Internet in the 80s and 90s, it is hard to imagine the world without it. It is one of the primary information sources nowadays. Almost everybody uses the Internet to find information (Jansen & Spink, 2006; Rouet, Ros, Goumi, Macedo-Rouet, & Dinet, 2011). Sometimes, just a simple search suffices to get the information that is needed, e.g., when looking for a fact like the height of the Eiffel tower. Whereas in other more multifaceted situations (e.g., searching information on controversial topics) a more complex search process is needed to find all required information, such as the causes of obesity. Irrespective of the type of the search task, people often find information they may want to use later (Jones, 2007). As a consequence, they need to re-access this information they found, and the question arises how to keep interesting information sources for later use. On top of that, usually people want to keep information about different topics, such as when collection information about a prolonged illness, for a hobby, in professional settings, or for a school assignment. When the information is stored – usually in the form of bookmarks – soon the need to organize the different sources becomes evident.

In this study we investigated whether providing support for organizing information using bookmarks has an influence on the quality of selecting and storing information. Before we will elaborate on the need for such support, we will review the literature on how information on the Internet can be re-accessed, how bookmarks can function as a personal library and how information can be organized and classified.

Re-accessing information on the Internet

Especially when information is collected over time and/or not used immediately, it becomes important to know how to re-access previously found information. Tauscher and Greenberg (1997) already showed that a majority of webpage visits are revisits. This was corroborated by Teevan, Adar, Jones and Potts (2007) who also showed that a large part of search queries were attempts to re-find information. As school assignment regularly span several weeks, this issue becomes a concern for students as well. In such situations students can rely on two strategies: search again when they need the information or save the information for later use. The strategy of searching again has the advantage that people do not have to think about how information has to be stored. However, when trying to re-access previously found information though searching again students are confronted with several barriers. First, the earlier used search terms are easily forgotten (Aula, Jhaveri & Käki, 2005; Teevan, 2008). When other terms are used to find a specific source again, it could be that this source will not show up in the search results again. Second, the specific source is not available anymore (Capra III & Pérez-Quinones, 2005). This could be due to a paywall or that the source is not online anymore. Third, search algorithms change over time (Marshall & Bly, 2005; Selberg & Etzioni, 2000). As a result, the source will show up in another place in the search results or even not show up at all. Fourth, the specific source has been changed and is therefore less recognizable (Adar, Teevan, Dumais, & Elsas, 2009). All those barriers result in a decreased ability to re-access previously found information by searching again.

Another and more optimal strategy is to save the found information. Saving information is not subject to changes in search algorithms, the recall of used search keywords, the availability of information (supposing a copy is made), or changes in the state in which the information was previously found (again, supposing a copy is made). Several previous studies (e.g., Aula, Jhaveri and Käki, 2005; Alhenshiri, 2013) describe several strategies used by people for saving information for later use such as: saving documents as a file, printing, add bookmarks to favorites, email the url, or writing down the url's. Of these strategies, the bookmarking system is the only tool that is specifically designed for the task of storing and organizing information from the Internet. A bookmark is a record of the address of a website, file, or other data made to enable quick access in future (Bookmark, n.d.). The concept of bookmarks was first introduced in the Mosaic browser in 1993 (Andreesen, 1993) and quickly gained popularity among users (Abrams, Baecker & Chignell, 1998; Pitkow & Kehoe, 1996). It is also the most used strategy for facilitating re-access (Bruce, Jones & Dumais, 2004). Over time such a collection of bookmarks will grow bigger and bigger (Abrams et al., 1998).

Bookmarks as a personal library

The main reason for bookmarking is to make sure that found information can be accessed easily in the future (Abrams et al., 1998). However, Abrams et al (1998) also indicate that to reap the benefits of such a bookmark collection, some sort of organizational structure for these bookmarks is necessary. However, people often do not structure their bookmark collection (Abrams et al., 1998; Stadtler & Bromme, 2008). Abrams et al (1998) posed the question of how to organize collections of many hundreds or even thousands of bookmarks already when the Internet was still in its infancy and thus much smaller than nowadays.

As the number of bookmarks grew over time, people started to organize their bookmarks to fight entropy and maintain accessibility (Abrams et al., 1998). You can structure bookmarks in folders, such as any other data on the computer. However, using folder bookmarks in the browser is inadequate (Heckner, Heileman & Wolff, 2009; Shen & Prior, 2013): (1) organizing bookmarks revolves mainly around the use of folders which can become difficult to manage when the number of folders and the levels in the folder hierarchy increases, (2) it is often difficult to name the folders, especially when people lack an existing knowledge structure about the subject, and (3) limited contextual information is available as mostly only the bookmark and folder names are available. As Shen and Prior (2013) report that although most people bookmark web pages regularly, they often also have problems finding back a recently visited web page. Thus organizing these bookmarks is important to keep them accessible. However, according to Jones (2007), the act of storing a bookmark is difficult and error-prone. The decision to bookmark a webpage elicits several implicit questions, e.g., Is the webpage useful? Should it be kept? Where? In what form? Often users develop a set of rules by which they organize their information (e.g., Whittaker & Sidner, 1996; Wash & Rader, 2007). Unfortunately, people often have a hard time maintaining a consistent organizing system over time (Golder & Huberman, 2006; Wash & Rader, 2007). Thus, strategies for organizing bookmarks are needed.

Two different ways of organizing bookmarks: hierarchical or tagging

When storing and organizing information sources on the Internet, one can essentially choose between two approaches of organizing information: a hierarchical or a tagging approach. In a hierarchical approach a user organizes bookmarks in a folder structure where folders can contain subfolders, which results in a tree-like structure. In the tagging approach bookmarks are not divided across folders, but are given labels. The tagging approach allows for assigning multiple tags to a bookmark and also one tag to different bookmarks. The bookmarking service Delicious™ introduced this approach to bookmarking in 2003 (“Social bookmarking”, 2015). Other examples of the tagging approach are Pinboard (bookmarking websites), Youtube (video), Flickr (photo’s), Connotea (science) or Last.fm (music).

Historically the dominant approach of organizing information is hierarchical because of the physical characteristics of information storing in offices and libraries (Wright, 2007). For example, although a book can be classified in several categories, a library can store this book in only one place (unless several copies are kept). Most studies take only a hierarchical approach into account (e.g., Chen & Dumais, 2000; De Vries, Van der Meij, & Lazonder, 2008; Stadtler & Bromme, 2008). However, several researchers claim that tagging systems can overcome several of the shortcomings of the traditional bookmarks in browsers (Golder & Huberman, 2006; Heckner at al., 2009; Wash & Rader, 2007). First, although the tagging approach might be more laborious, in the digital era it has become a viable solution for organizing information sources because one information object can now be assigned to several categories, thus alleviating classification problems. Second, by assigning multiple tags to a bookmark, a user is better able to add contextual information.

Unfortunately, there has been little research among users into the differences between those two approaches. Civan, Jones, Klasnja and Bruce (2008) compared two existing products (Hotmail, hierarchical & Gmail, tagging) with ten participants. Each participant got two assignments: for one assignment participants stored the information in folders and for the other in tags. Civan, et al (2008) interviewed participants afterwards about their experiences. They found several differences. On the one hand, organizing with tags requires more physical effort because regularly several tags were assigned to an information source which requires more clicks for the users. Moreover, hierarchical organizing requires a higher cognitive effort because the participants found it difficult to decide in which folder a specific information source had to be stored. On the other hand, most participants found that tags are a better and more flexible way for searching in the stored information. Not only the greater flexibility was appreciated, but the descriptive characteristics of tags were valued as well. According to the participants, tags give an indication about the content of the source. Besides that, participant also found it also easier to jump to related concepts. A study by Voit, Andrews and Slany (2012) involved both storing and re-finding information with either a folder structure or a tagging system. In both conditions the participants had to create the folder structure or the tags themselves. After a break, the participants had to re-find a set of six test items. They concluded that there are no significant differences between folder structures and tagging with regard to storing and re-finding behavior. Research by Bergman, Gradovitch, Bar-Ilan and Beyth-Marom (2013b) asking 168 students about their preferences for either folders or tags, revealed several interesting findings. Although they see a favorable attitude towards the use of tagging for managing information, actual behavior shows that folders are preferred over tags. They indicate that familiarity with using folders is the driver for this discrepancy. Also, perceived simplicity is a possible explanation for this preference. Pak, Pautz and Iden (2007) compared performance and subjective workload between using folders and using tags. In the two experiments they carried out, they found mixed results. On the one hand, participants experienced more frustration when organizing with tags, while the results with regard to actual performance were mixed. On the other hand, a tagging system leads to better re-finding requiring less clicks to get to the desired result. This might indicate that using tags is a viable alternative to the use of folders for organizing information. Especially for large amounts of information tags might unfold their potential with regard to easier accessibility (Walhout et al., 2015).

Prior knowledge and the need for support

Unfortunately, as shown by Stadtler and Bromme (2008), saving information for later re-access is accompanied with its own problems. The participants in their study did not organize the found information in a proper way and they neither structured the information according to straightforward principles, which makes retrieving previously found information problematic. A possible reason for this might be the lack of domain knowledge, which hinders the construction of such a classification scheme (Walhout, Oomen, Jarodzka & Brand-Gruwel, 2017; White, Dumais & Teevan, 2009). To overcome this lack of knowledge, especially for novices, providing support can be beneficial (Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga, 2007). Novices cannot rely on an existing knowledge structure because they have a low level of prior knowledge about the subject at hand. It is therefore difficult for them to come up with a classification scheme to classify the found information. Supporting them with a given classification scheme can function as a replacement for such an already existing knowledge structure. On the contrary, students with a higher level of prior knowledge might be hindered by such support as it can contradict with their existing knowledge structure.

People develop their own information classification scheme based on their needs and prior knowledge (Bergman et al, 2013a). This matches with findings from cognitive science. When judging the relevancy of new information, this new information is connected to existing knowledge, and organized into already existing schemata (Rogers & Swan, 2004). However, when there is no or not much existing knowledge, there are also no schemata (Kalyuga et al., 2003; Kalyuga, 2007). Hence, people with low or no prior knowledge on a topic might need support to develop a classification scheme. Research also suggests that providing any form of support might be beneficial for people with low prior knowledge, however, it might be counterproductive for people with high prior knowledge as suggested by the expertise reversal effect (Kalyuga et al., 2003; Kalyuga, 2007). For people with low prior knowledge, support in the form of a given classification scheme might function as a replacement scheme for the knowledge structures that are used by people with high prior knowledge. On the other hand, individuals with higher levels of prior knowledge rely on their existing prior knowledge structure when solving complex cognitive problems like information problems. Providing them with a classification scheme might therefore lead to an increased cognitive burden because they have to match their own scheme with the provided one.

Thus, one may assume that supporting low prior knowledge participants with a classification scheme will result in a better selection and organizing of information and that they will be better able to connect new information with existing knowledge. This is indeed supported by the research of De Vries, Van der Meij and Lazonder (2008). In their study of search behavior they gave pupils a structured set of information sources. In their first experiment the sources were structured in only five broad categories, while in their second experiment the sources were structured with a detailed classification scheme. Compared to the first experiment, the number of correct answers increased considerably in the second experiment. Thus, supporting students with organizing information sources helped them selecting relevant information.

A study of Stadtler and Bromme (2008), although not about bookmarking, gives another clue that giving support is beneficial for novices. In their research participants with little medical knowledge were asked to complete a search task about cholesterol. With the tool *met.a.ware* participants could store information. In this tool a classification scheme about cholesterol was incorporated. Stadtler and Bromme (2008) found that the use of the supporting tool *met.a.ware* had a positive effect on the organization of notes and that the tool had a positive effect on knowledge acquisition about cholesterol. Moreover, the control groups showed hardly any type of structuring of their saved sources. These results suggest that providing support can help in organizing and possibly also in re-accessing a collection of information sources.

THIS STUDY

This study focuses on the interaction between bookmarking behavior, support, and prior knowledge. The central research question for this study is: How is bookmarking behavior influenced by support with a classification scheme and by support with different types of organizers? To answer this question, the following sub-questions will be addressed: 1) Does providing support influence bookmarking behavior? Two types of support will be distinguished in this: content support with a classification schema and tool support with either a tagging or a hierarchical system. 2) Does prior knowledge influence bookmarking behavior? For this question, also a possible interaction between prior knowledge on the one hand and type of classification support and type of organizer on the other hand, will be studied.

Bookmarking behavior is in this study expressed by several measurements, such as, number of bookmarks (cf. Abrams et al., 1998; Boardman & Sasse, 2004; Wash & Rader, 2007), quality of the bookmarked pages and diversity of the used classifications (cf. Boardman & Sasse, 2004; Wash & Rader, 2007). Support is given along two dimensions: type of classification support (no classification scheme versus a given classification scheme) and type of organizer (hierarchical versus tagging). Prior knowledge is taken into account to see whether this interacts with bookmarking behavior.

METHOD

Participants

A total of 95 ninth grade students (37 male, 58 female; M_{age} 14.49 years, $SD_{age} = 0.6$) in the highest level of secondary education in the south of The Netherlands took part in this study. Prior to the experiment, both the students and their parents received an explanation of the study (students during the lesson, parents via a letter of consent).

Material

Task. To ensure both experimental control and ecological validity, a simulated task like the one used in this study should meet several requirements. The task needs to have content to which the participants can relate, it should provide enough imaginative context and it should allow for multidimensional and dynamic relevance judgments (Borlund, 2000 & 2003). The multi-dimensionality stems from the fact that information is not only judged on its usefulness, but also on dimensions such as quality, topicality, or the origin of the information. Dynamic relevance refers to the changing importance of relevance criteria during the information solving problem process. This multi-dimensionality was reflected in the evaluation protocol used to determine the quality of the search results.

The task was constructed in cooperation with teachers to ensure that the task domain and workload was comparable to what is common in education and this group of students. To ensure enough interest in this topic, a pilot was conducted. Several groups of students were asked to indicate which subject from a list of five they would like to know more about. The subject of 'obesity' was named most frequently.

Hence, a task on the topic 'obesity' was constructed. It is a topic that students are vaguely familiar with, but it is no part of their regular curriculum. Participants were asked to find an answer to three questions: (1) what is obesity, (2) what are the causes of obesity and (3) what are the consequences of having obesity. From the Google-like search results page the participants had to locate and store the relevant webpages with a bookmarking tool. At the end the participants had to indicate which of the bookmarks were the three most important ones to answer the three questions.

Search result pages. A Google-like search result page with 25 preselected search results was constructed. The order of the search results was varied and five different search results pages were constructed. The search function and all links except the search results were disabled. When constructing this list of websites, we ensured that a heterogeneous set of sources was included.

Bookmarking support. For this experiment a bookmarking tool was developed that enabled four different conditions: hierarchical structuring with no classification support, hierarchical structuring with classification support, tag cloud structuring with no classification support, and tag cloud structuring with classification support. To bookmark a webpage, the participants had to click on a button in the browser (see Figure 1) after which a pop-up appeared (see Figure 2) in which they could specify how to save the bookmark. The options in the pop-up depended on the conditions to which the participant was assigned. In the web-application the participants could star their bookmarks, thus indicating the importance of the bookmark.

Figure 1: Bookmark button in browser window

Figure 2: Bookmark pop-up

Software. BB Flashback from Blueberry software was used to capture the activity on the computers. Data processing and data analysis was done with the R language for statistical computing (R Core Team, 2017).

Measurements

For each of the research questions quantitative as well as qualitative measurements were taken into account. Bookmarking behavior was operationalized by the *number of bookmarked webpages,* the *quality of the bookmarked webpages,* the *diversity of the used classifications*, and a *similarity score*.

The *number of bookmarked pages* were counted. To determine the *quality of the bookmarked webpages*, two independent raters scored each of the 25 webpages on a scale of 1 to 10 on several aspects: connection to the task, appearance of the webpage, target audience, reliability of the organization or person behind the website, the use of references to other sources, and whether information about the author was given. Based on the scores on each aspect an aggregated score for each website was calculated. The interrater reliability, expressed as Pearson's correlation coefficient, showed a significant correlation (r = .89, p < 0.01). Close examination of the aggregated scores by the two raters revealed several differences in score which were larger than 1.5. These webpages were re-evaluated by the two raters. After this re-evaluation, the interrater reliability increased slightly (r = .96, p < 0.01). Thus the interrater reliability can be considered high. For each website the average of these scores was calculated and assigned to the bookmarks. For each participant a *quality score* was calculated by taking the mean of the scores of the bookmarked webpages. Furthermore, the participants had to indicate a *top*

3 in their bookmarks, for which a separate *quality score* was calculated, thus resulting in a *top 3 quality score*.

The diversity of the used classifications was measured as the number of unique classifications used during the task.

Another way of looking at the effect of classification support, is measuring whether participants use classifications in the two no-classification support groups and to which extent this differs from the classifications the other two groups received. To measure whether no classification support would lead to the use of other classifications, a *similarity score* was calculated for each participant in the two no-classification support groups. This *similarity score* expressed the resemblance of the used classification and the corresponding given classifications (i.e., the classifications of the users in the tagging no support group were compared with the given classification for the tagging classification support group).

To measure *prior knowledge* of participants, they were asked to create a mind map of what they knew about the topic of obesity prior to the experiment. First, the number of nodes in the mind maps was calculated as the *prior knowledge usage score.* Second, the words used by the participants in the mind maps were matched with the used classifications. Thus, the *prior knowledge usage score* expresses the number of words from the prior knowledge of the participants that were used as a classification.

Design

To study the influence of support on bookmarking behavior in solving information problems, a 2 x 2 factorial design was used with the factors type of organizer (hierarchical structure vs. tag cloud) and classification support (no support vs. given classification scheme) as between subject factors. The dependent variables were bookmarking behavior measurements (see above). Participants were divided across four groups: hierarchical structuring with no classification support (n=27), hierarchical structuring with classification support (n=24), tag cloud structuring with no classification support (n=19), and tag cloud structuring with classification support (n=25).

Procedure

For each participant, the demographic data (age and sex) and prior knowledge were collected several weeks before prior to the experiment. The experiment was conducted in group sessions, one sessions for each condition. Each session started with a detailed explanation of the procedure. Participants were given 30 minutes to complete the task in which they had to collect at least three information sources. The sessions were recorded with BB Flashback. The total duration of the sessions was approximately 50 minutes (including instructions). After the instruction, the participants performed the search task. When they had completed the task, they could stop and save the screen recording.

RESULTS

Parts of the data were not normally distributed, thus, non-parametric statistical tests (Wilcoxon rank sum test and Kruskal-Wallis test for between subject factors) were used to analyze these data when transformation did not result in normally distributed data. For the normally distributed variables, parametric tests were used. All relevant means and standard deviations are summarized in Table 1.

Table 1 Means and Standard Deviations Bookmark Data

	Classification support		No classification support	
	Hierarchical	Tag cloud	Hierarchical	Tag cloud
N	24	24	27	19
Number of bookmarks	10.50 (4.03)	7.63 (3.74)	8.44 (2.90)	9.42 (3.99)
Quality score	5.74 (0.39)	5.70 (0.49)	5.71 (0.46)	5.84 (0.44)
Top 3 quality score	5.67 (0.72)	5.97 (0.49)	5.64 (0.57)	5.85 (0.56)
Number of unique classifications	3.42 (1.77)	7.29 (2.77)	3.93 (2.23)	6.42 (3.91)
Prior knowledge usage score	0.54 (0.51)	0.64 (0.70)	0.07 (0.27)	0.37 (0.50)

Influence of support on bookmarking behavior

A two-way independent ANOVA revealed a significant effect for the interaction between type of organizer and classification support ($F(1, 90) = 6.43$, $p = 0.01$, $\eta^2 = 0.07$) on the *number of bookmarked webpages*. The participants in the hierarchical organizer classification support group bookmarked more webpages than the participants in the tag cloud organizer classification support group. No significant differences were found for the main effects organizer ($F(1, 90) = 1.56$, $p > 0.05$) and classification support ($F(1, 90) = 6.43$, $p > 0.05$).

There was no difference in the *quality of all bookmarked webpages* with regard to type of organizer ($F(1, 90) = 0.63$, $p > 0.05$), type of classification support ($F(1, 90) = 0.59$, $p > 0.05$) or an interaction between these two factors ($F(1, 90) = 0.38$, $p > 0.05$) when looking at all bookmarked webpages.

When looking at the *top 3 quality score*, there was a significant difference with regard to the type of organizer ($F(1, 90) = 4.16$, $p = 0.04$, $\eta^2 = 0.04$). The participants in the tag cloud group selected webpages of higher quality than the participants in the hierarchical organizer group. As the effect size is rather small, the effect of type of organizer on the quality of the bookmarked webpages is not very distinct.

With regard to the *diversity of the used classifications*, the analysis was split in two parts: a separate analysis for each organizing system. This is based on the fact that inherent to the nature of the classification systems the participants in the tag cloud condition could use more than one classifier for each bookmark. An independent t-tests showed no significant differences in the *number of unique classifications* between the groups getting classification support and the groups getting no support for both the tag cloud organizer ($t(32) = -0.82$, $p > 0.05$) as well as the hierarchical organizer ($t(48) = 0.91$, $p > 0.05$).

The *similarity of the self-made classifications* in both no-classification support groups with the respective given classifications was significantly higher for the participants in the tag cloud organizer no-classification support group than for the participants in the hierarchical organizer no-classification support group (W = 198, p < .027, r = □0.227).

Influence of prior knowledge on bookmarking behavior

Regression analysis did not show an influence of the level of prior knowledge on the quality of the selected bookmarks ($R^2 = .02$, $F(1, 90) = 1.41$, $p > 0.05$), the quality of the top 3 bookmarks ($R^2 = .001$, $F(1, 89) = 0.09$, $p > 0.05$), the number of bookmarks ($R^2 = .001$, $F(1, 90) = 0.10$, $p > 0.05$), or the used classifications ($R^2 = .001$, $F(1, 90) = 0.06$, $p > 0.05$).

The *prior knowledge usage score* was significantly higher for the participants who were given a classification scheme compared to the participants who had no support of a classification schema (W = 746, p < .001, r = 0.343). A Kruskal-Wallis test revealed a significant difference for the interaction between type of organizer and type of classification support ($H(3) = 14.257$, $p = .003$). The participants in the hierarchical organizer no-classification support group had a lower score than the participants in the tag cloud organizer classification support group.

DISCUSSION

People often want to store found information on the Internet for later use (Jones, 2007). Bookmarks are a way to store links to webpages with interesting information and help in re-accessing previously found information easily in the future (Abrams et al., 1998). To support people in managing their bookmarks, several so-called social bookmarking services were developed, e.g.: itList, Webtagger, Backflip, Clickmarks, etc. ("Social bookmarking", 2015). Like bookmarks that are stored in the browser, the bookmarks which were stored with these services had to be organized in folders. However, when such a collection of bookmarks grows bigger, the need for organizing this bookmark collection arises (Abrams et al., 1998). Basically two systems for organizing can be distinguished: a hierarchical (using folders) or a tagging approach (using keywords, which was introduced by Delicious in 2003) can be used. Moreover, prior knowledge about a domain enhances the ability to organize found information (Kalyuga et al., 2003; Kalyuga 2007; Špiranec & Ivanjko, 2013). Therefore, to provide support it is not only important which type of organizing system works best, but also whether predefined (or suggested) schemata have an effect on this as well.

The purpose of this study was to gain insight in the influence of classification support (no vs. given classification) and support with an organizing system (hierarchical vs. tag cloud) on bookmarking behavior and in what way this interacts with prior knowledge. Bookmarking behavior was operationalized by the number of bookmarks, the quality of the bookmarks, the diversity of used classifications and the similarity of the used classifications.

The results with regard to bookmarking behavior showed only a weak interaction effect of the two types of support on the number of bookmarks. Furthermore, students receiving support with a classification scheme used more classifications that were already present in their prior knowledge. This corroborates the findings of earlier outcomes that indicated that giving support might be beneficial for novices (Kalyuga et al., 2003; Kalyuga 2007; Špiranec & Ivanjko, 2013; Rader & Wash, 2008; Stadtler & Bromme, 2008).

Giving support with either a tool or a classification scheme did not have any effect on the *quality of all bookmarked webpages*, the quality of the top three of bookmarked pages, the *diversity of the used classifications* and the *similarity of the self-made classifications*. Based on previous research (Kalyuga et al., 2003; Kalyuga 2007; Rogers & Swan, 2004; Špiranec & Ivanjko, 2013; Stadtler & Bromme, 2008) we expected to see differences on these measurements.

A possible explanation for these inconclusive results could be the low familiarity of participants with using a tagging system to organize information. This gave the participants in the hierarchical conditions an advantage because they could use a system they were familiar with. Although people show a preference when asked, they fall back to what they already know and are used to in practice (Bergman et al., 2013b). Of the previous studies only the study of Bergman et al (2013a) used a short familiarization period for the tagging approach before the participants were given a free choice of approach. As also the previous studies presented mixed results, it remains therefore unclear whether using a hierarchical system or a tagging system is better.

In our study we did not use such a short familiarization period, which can be seen as a limitation to our study because one might argue that not using such a familiarization period leads to a disadvantage for the less know approach. Another limitation is that this study focused on the observed behavior in the bookmarking process. However, this bookmarking process is part of a larger entirety: the search process. To improve the ecological validity in the future, it is important to embed the different conditions of this study in a more complete search process. Finally, the focus on the observed behavior leaves the opportunity open to study the motivations behind the choice of tags. This can be done using a think aloud procedure or cued retrospective report (Brand-Gruwel, Kammerer, Meeuwen & Van Gog, 2017).

CONCLUDING REMARKS

The most important conclusion is that, despite the low familiarity with tagging systems, using a hierarchical system (i.e., folders) is not necessarily a better approach than using a tagging system to organize previously found information. For future research it might therefore be worthwhile to investigate the differences between the hierarchical and the tagging approaches while taking the low familiarity of tagging into account. To do that, participants need to be either experienced with the tagging approach or need to get an extensive period to gain such experience. Lee, Goh, Razikin and Chua (2009) already showed that high familiarity with the concept of tagging leads to better performance. Whether this also leads to a better performance in comparison to more traditional folder-based systems, needs to be investigated. Furthermore, different uses of tags might influence the outcome as well. Basile, Peroni, Tamburini and Vitali (2015) make a rather general distinction between topical (describing the contents of the source) and non-topical (describing non-content aspects of the information) use of tags. Another, similar, classification of the types of tags is the distinction between content tags, attitude tags, or self-reference tags (Melenhorst & Van Setten, 2007; Arolas & Ladrón-de-Guevar, 2012) in which the content tags are similar to the topical tags and the other two types are similar to the non-topical tags of Basile et al (2015). In a study by Fastrez and Jacques (2015) tags were exclusively used to describe the contents of a source. This finding suggests that tags might be better suited for topical or content use.

The fact that in our study participants in the tag-cloud-no-organizer-support group produced classifications more similar to their prior knowledge than participants in the hierarchical-organizer-no-support group in combination with the known effects of higher prior knowledge, might lead to the expectation that people with higher prior knowledge will make use of topical tags more often than people with a low level of prior knowledge.

Chapter 5

Organizing bookmarks with tags: Effects of prior knowledge and given classification support

Abstract

When students want to store found information, the question of how to organize the growing collection of sources soon arises. However, novices do not have a well-developed cognitive schema and thus might have difficulties in organizing sources properly. Supporting them with a classification schema can help in organizing information. This study investigates the effect of providing support with an expert defined classification schema on experienced task difficulty, tagging behavior and processing of a bookmark tool. The main conclusion is that providing support with a classification schema to organize information influences their bookmarking behavior. As a result, their information storing and organizing behavior might be improved as well. Providing support has a positive influence on the number of tags used to store a bookmark and that giving support elicits the use of concepts which were already part of the prior knowledge of the participants. Furthermore, the use of less unique tags indicates a more focused use of tags to organize the information. This might indicate that providing support leads to a less cluttered classification of the collected information. No evidence of an interaction with prior knowledge was found.

In contemporary educational settings, students often acquire knowledge on a subject by collecting information from the Internet and putting that information into context. To do so, students should formulate their information needs, find information sources, distil and organize the relevant pieces from these sources and combine this information from several sources in a meaningful way (Brand-Gruwel, Wopereis & Walraven, 2009). This is referred to as information problem solving (IPS) (e.g.,: Brand-Gruwel et al, 2009; Eisenberg & Berkowitz, 1990; Wolf, Brush, & Saye, 2003). To complete IPS tasks, student must have at least some prior knowledge on the subject. Otherwise, they engage in a clueless search for information and, as a consequence, hardly learn anything (Clark, Kirschner & Sweller, 2012).

Previous research has shown that students often experience problems when trying to solve IPS tasks (e.g.,: Brand-Gruwel, Wopereis, & Vermetten, 2005; Large & Beheshti, 2000; Walraven, Brand-Gruwel, & Boshuizen, 2008). These problems are not only limited to students in educational settings. People in all age groups and outside formal educational settings often lack the skills to solve IPS tasks (Walraven et al, 2008). Mastering these skills thus needs extensive instruction. There is a substantive body of research for several of the IPS subskills (e.g.,: Eisenberg & Berkowitz, 1990; Salmerón, Kammerer, & García-Carrión, 2013; Walraven, Brand-Gruwel & Boshuizen, 2009) and how instruction can help students to develop these subskills (e.g.,: Alhenshiri, Watters, Shepherd & Duffy, 2012; De Vries, Van der Meij & Lazonder, 2008). Searching the Internet, storing and organizing the found information for later use has received relatively little attention in research. The existing research on that topic is mostly from the personal information management perspective while the educational perspective is often missing. The subskills of storing and organizing of the found information is especially important when information is not processed immediately and needs to be re-accessed at a later time. Among the several possible strategies to enable easy re-access of previously found information for later use, the bookmarking system is specifically designed for storing and organizing information from the Internet (Abrams, Baecker, & Chignell, 1998; Andreesen, 1993; Aula, Jhaveri & Käki, 2005; Alhenshiri, 2013).

Research reveals that people often do not properly structure the information they find (Abrams et al., 1998; Stadtler & Bromme, 2008). This can cause problems when people want to re-access that information on a later time. Abrams et al (1998) already posed the question how to organize such a collection of information. Organizing found information using keywords and a classification scheme requires at least some prior knowledge. Although people often develop their own classification scheme, they also have difficulties maintaining a consistent system over time (Bergman et al, 2013a; Golder & Huberman, 2006; Wash & Rader, 2007; Whittaker & Sidner, 1996). So, to organize found information properly, it is necessary that the students can rely on an already existing cognitive schema about the subject of the task (Walhout, Oomen, Jarodzka & Brand-Gruwel, 2017; White, Dumais & Teevan, 2009; Wildemuth, 2004). Unfortunately, in an educational setting this is often not the case. Providing students with classification support could help them to better organize the found information.

In this study, we investigate the sub processes of storing and organizing found information and whether providing classification support has an effect on this. First, we will review the literature on providing support and its relation with prior knowledge. Next, we will discuss how such support might be constructed.

Prior knowledge and providing support

Prior knowledge on a subject aids successful IPS (Clark et al., 2012; White et al., 2009). Studies on the interaction between IPS and expertise show that experts and novices differ significantly on key aspects of the search process like query formulation, website selection, and task effectiveness (e.g., Brand-Gruwel et al., 2017; Chen & Macredie, 2010; Monchaux et al., 2015; Lee & Pang, 2017; White et al., 2009). Prior knowledge is also key for developing cognitive schemata on a subject (Bergman, et al., 2013; Kirschner et al., 2006). Newly found information is immediately compared with what is already known and thus judged in the context of the already existing cognitive schemata (Kirschner et al., 2006; Rogers & Swan, 2004). However, when prior knowledge is very limited or even completely lacking, internal cognitive schemata are either poorly developed or completely non-existent (Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga, 2007). Novices can thus not rely on an existing knowledge structure which makes it difficult for them to classify and organize the new information (Clark et al., 2012; Rogers & Swan, 2004; Walhout, et al., submitted). From cognitive load theory, we know that when novice learners encounter new information they continuously have to assess the current state of the problem solving process with the goal of the problem (Sweller, 1988). This places a high burden on working memory, which gets easily overloaded in such situations. Consequently, working memory is severely limited from adding the new information to the long-term memory (Clark et al., 2012; Kalyuga et al., 2003; Kalyuga, 2007; Kirschner et al., 2006). Prior knowledge about a domain enhances the ability to organize found information (Kalyuga et al., 2003; Kalyuga 2007; Špiranec & Ivanjko, 2013). Therefore, a lack of prior knowledge and its accompanying schemata result in difficulties for novices to classify newly found information. In such situations, providing support to novices lowers the burden on their working memory and, as a result, improves the transition of information between working memory and long-term memory. Thus, enhancing comprehension and improving task success. However, providing support is not beneficial for everyone. Support might even be counterproductive to learners with high prior knowledge. The support is in such cases often redundant and consequently distracts the working memory from retrieving information from long-term memory (Clark et al., 2012; Kalyuga et al., 2003; Kalyuga, 2007; Kirschner et al., 2006).

The form of support in relation to prior knowledge

In the previous section, we discussed the relation between prior knowledge and the need for support when organizing found information. It became evident that for non-experts at least some guidance is necessary. The level of support that is needed depends on the level of prior knowledge (Van Merriënboer & Kirschner, 2012). Within the scope of this article the question arises how support can be designed for giving students guidance during the sub process of storing and organizing information found on the Internet. In this paragraph, we will review previous research for clues about the form of the support that can be given.

As we have previously seen, students with low prior knowledge on a subject have no cognitive schema and thus organize their information sources poorly. Consequently, one can thus hypothesize that an obvious kind of support would then be to provide students with a classification scheme which they can use to organize the found information (Walhout et al, submitted). This might result in in a better selection and organizing of information. De Vries et al (2008) carried out two experiments on search behavior; one in which they provided participants with a small set of broad categories and another one in which they provided participants with a detailed classification scheme. In the experiment with the detailed classification scheme, participants gave considerably more correct answers than in the experiment with the broad categories. From this study, it can be concluded that providing detailed support is beneficial for novices when carrying out a search task. These findings are in line with the proposed level of guidance by Van Merriënboer and Kirschner (2012): novices should get full guidance in their learning tasks.

The study of Stadtler and Bromme (2008) also indicates that novices benefit from receiving support. In their study, they asked participants with little medical knowledge to complete a search task on the topic of cholesterol. The support tool the participants could use to store information included a classification scheme on cholesterol. Compared to the control group, the participants who used the support tool organized their information better and also acquired a better understanding of the topic of cholesterol.

The findings of De Vries et al (2008) and Stadtler and Bromme (2008) are corroborated by our previous study (Walhout et al, submitted) in which we found that providing students with a classification scheme resulted in the use of classifications that were already present in their prior knowledge. This might indicate that providing students with a classification scheme helps them to elicit tacit knowledge of which they were not aware spontaneously.

Type of support

The previous section has shown that providing students with support via a classification scheme, helps them to organize information and supports their learning. As described earlier, the bookmarking system is specifically designed to collect, store and organize information on the Internet. Traditionally, information is organized in hierarchical structures (Wright, 2007). However, organizing bookmarks in a hierarchical structure with folders might become complicated when a collection grows (Heckner, Heileman & Wolff, 2009; Shen & Prior, 2013). Previous studies have highlighted a major shortcoming of hierarchical structures: it forces students to choose one of several possibilities for classifying information (Bergman et al., 2013b; Dourish et al., 2000). This forced choice can increase cognitive load because students have to decide which category is the best choice (Dumais & Landauer, 1983; Lansdale, 1988; Malone, 1983). Furthermore, when making such a decision, students also have to think about how they are going to use the information in the future (Bruce, 2005; Kidd, 1994).

In contrast to the hierarchical folder structure, tagging allows using several labels for one information source. The tagging approach was first introduced for bookmarking by Delicious™ in 2003 ("Social bookmarking", 2015). Based on their research, several researchers think that a tagging approach might address the drawbacks of a hierarchical system (Golder & Huberman, 2006; Heckner at al., 2009; Wash & Rader, 2007). When tagging, no classification choice needs to be made, thus lessening the cognitive load of classifying information. Moreover, because multiple tags can be assigned to one source, it becomes easier to add contextual information and makes sure that there are several ways to retrieve the information later. Several studies have compared the hierarchical approach and the tagging approach with mixed results (Bergman et al., 2013b; Civan, et al., 2008; Pak, Pautz & Iden, 2007; Voit, Andrews & Slany, 2012; Walhout et al., 2015; Walhout et al., submitted). The participants in the study of Bergman et al (2013b) had a positive attitude towards the use of tags, but their actual behavior showed a favor for folders. Tags are more descriptive, require more physical actions, but result in better and more flexible retrieval, while hierarchical folders are associated with higher cognitive load (Civan et al., 2008). Pak, Pautz, and Iden (2007) found that participants were more frustrated when organizing with tags and their actual performance was mixed. On the other hand, tagging led to more efficient retrieval of previously stored information. This outcome differs with that of Voit, Andrews and Slany (2012) who reported no difference between folder structures and tagging with regard to storage and retrieval. In an earlier study, we found no difference in task performance when navigating a collection of information sources with either a hierarchical menu or a tag cloud (Walhout et al., 2015). Although there were no differences in task performance, participants in the tagging condition showed a more focused selection of sources. This was corroborated with a follow-up study in which we also did not find significant differences between using a hierarchical system or a tagging system (Walhout et al., submitted). Although the participants in both studies were not familiar with the use of tagging, they did not perform worse. As Lee et al. (2009) showed, a higher familiarity with the concept of tagging improves performance. However, we do not know yet how the perceptual processing of such systems work. The visual processing can be studied with eye-tracking methods (cf.,: Walhout et al., 2015 & Walhout et al., 2017).

THIS STUDY

The central research question for this study is: What is the effect of providing a classification scheme for bookmarking while searching the Internet considering the effect of students' prior knowledge on experienced task difficulty, tagging behavior, and the way students use the given bookmark tool?

Prior knowledge is defined as (a) the breadth of knowledge by the number of keywords students produce on the topic and (b) the quality of knowledge established by comparing it to an expert model. Well-developed cognitive schemata help students to sort and categorize information they find. Hence, we hypothesize that students with a well-developed cognitive scheme (high prior knowledge) will report low experienced task difficulty irrespective whether the student received on classification scheme for bookmarking or not (H1a). If the prior knowledge is low, then students will report lower experienced task difficulty in the given classification scheme condition and high experience task difficulty in the no-classification scheme condition (H1b).

With regard to tagging behavior, students with a higher prior knowledge will use more tags per bookmark in both conditions (H2a). They will also use more unique tags in both conditions (H2b). When the prior knowledge is low, then students will use more tags in the given classification scheme condition and less tags in the no-classification scheme condition (H2c). Students with low prior knowledge will use less unique tags in both conditions (H2d). Furthermore, students in the given classification scheme condition will use more tags which are similar to their prior knowledge (H2e).

Concerning the use of the bookmark tool, we expect that students with high prior knowledge will spend less time on the bookmarking tool in the given classification scheme condition and more time in the no-classification scheme condition (H3a). They will also need less time for the first interaction with the tool in the given classification scheme condition and more time in the no-classification scheme condition (H3b). When prior knowledge is low, students will spend more time on the bookmark tool in the given classification scheme condition then in the no-classification scheme condition (H3c). These students will need more time for the first interaction with the tool in the given classification scheme condition and less time in the no-classification scheme condition (H3b).

METHOD

Participants

In this experiment 68 school students (45 male and 23 female; M_{age} = 15.31 years, SD_{age} = 0.65) from the highest level of secondary education of in the south part of The Netherlands participated. All participants had experience with searching on the Internet. Participation was voluntarily and was rewarded with a gift certificate of 15 euro.

Design

In this study, we used a between-subjects design, whereby participants in the experimental condition received a given classification scheme while participants in the control condition received no such scheme. Participants were randomly assigned to one of the two conditions.

Material and apparatus

Task. Participants conducted a simulated exploratory search task on the Internet, because for such tasks they have to collect information from different sources, these tasks are open ended, rather general than specific, involve uncertainty, are associated with goals of learning / investigation, need multiple items to solve the task, and occur over time (Wildemuth & Freund, 2012). Moreover, simulated search tasks facilitate both experimental control as well as realism by describing a realistic information-requiring situation that motivates the participants to search for information (Bell & Ruthven, 2004; Borlund, 2000 & 2003; Borlund & Schneider, 2010).

The topic of the search task was 'climate change'. It is a topic that students are familiar with, but it is not a part of the regular curriculum. The task formulation was as follows:

> *Suppose you are a journalist for a large national newspaper (e.g., the "Volkskrant"). The editor in chief asks you to write a background article to investigate the question: Is mankind responsible for global warming? The aim is to reflect critically on the arguments in favor and against the influence of mankind on global warming. You have approximately 30 minutes to collect minimal twenty resources, which you can use to make an outline for such an article.*

During the experiment, participants searched via the Google® search engine. Apart from that, they could search the Internet without any restrictions. Students located and stored relevant information using a bookmark tool while searching and navigating through pages on the Internet. To ensure that the task content and workload were comparable to what is common in education, we constructed this task in cooperation with teachers.

Bookmarking tool. For this experiment a bookmark tool was developed that enabled two different ways of tagging: one condition with a classification scheme and one condition without. This bookmarking tool is a web application that resembles social bookmarking services like Delicious and Pinboard. To bookmark a webpage, participants clicked on a button in the browser (see Figure 1) after which a pop-up appeared (see Figure 2) in which they could specify how to save the bookmark. The options in the pop-up depended on the condition to which the participant was assigned. In the given-schema condition, participants could choose from the given tags suggestions area. They were not allowed to create new tags. The participants in the no-schema condition filled in their own tags in the 'tags' text field. In subsequent activations of the pop-up, the suggestions part was populated with the previously used tags (see Figure 3). In the web-application the participants could star their bookmarks, thus indicating the importance of the bookmark.

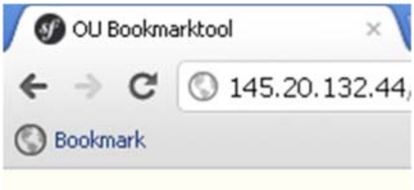

Figure 1: Bookmark button in browser window

Figure 2: Bookmark pop-up given schema

Figure 3: Bookmark pop-up no schema

Classification scheme. We used the Delphi method consulting several experts on the topic of global warming to create a classification scheme for the bookmark tool that provides an appropriate representation of the topic (Linstone & Turoff, 1975). The Delphi method is an often-used method to achieve consensus and aims to systematically collect and process the views or opinions of people (often experts) on a certain topic. The consensus on the topic under consideration is reached through a process of iterative dialog (Jones et al., 2015; Linstone & Turoff, 2011; Mankoff, Rode & Faste, 2013). Although there are many variants of the Delphi method, the key features are anonymity, iteration, controlled feedback, and aggregation (Mankoff et al., 2013):

Table 1 Features of the Delphi method (Mankoff et al., 2013)

Anonymity	Participants don't see each other and mostly also don't know each other.
Iteration	The same problem or question comes back in several rounds. This iteration is probably the most distinctive aspect of the Delphi method.
Controlled feedback	The results of each round / phase are summarized by the experimenter(s) and given back to the participants in the next round.
Aggregation	The results of each round are often, but not always, quantitatively aggregated.

Our approach with the Delphi method included all those key features where the aggregation was conducted qualitatively instead of quantitatively. The process used in this study consisted of five phases. First, we contacted potential (semi) experts through email to ask for their participation. These experts were (assistant) professors in the faculty of Natural Science (with a specialization in Sustainability and Environment) at the Open University of The Netherlands and teachers in secondary education who constructed lesson plans on this topic. Six experts agreed to participate; four (assistant) professors and two teachers. Second, these experts received an email with the question to describe the subject of global warming in approximately twenty keywords or keyword combinations. They were also asked to avoid technical terms as much as possible and use the laymen's equivalent instead. Third, the experimenter summarized these responses in a classification scheme of keywords. In the fourth phase, we presented this summarized collection of keywords to these experts for feedback. They indicated which keywords they did not agree on and by which other words they would replace those keywords. In the fifth phase, we adapted our classification scheme based on this feedback. This resulted in the adaptation of three keywords. The final classification scheme consisted of 19 keywords.

Equipment. Eye movements and logging data were recorded with a remote SMI RED 250 eye tracker with a temporal resolution of 250 Hz, which is integrated in a PC screen, and is operated with SMI iViewX and SMI ExperimentCenter software from the stimulus PC. During recording, we used the screen capture recording mode, so not only the eye movements, but the entire task performance process (including possible mouse and keyboard operations) was captured. Data processing and data analysis was executed with the R language for statistical computing (R Core Team, 2017).

Measures

Prior knowledge. Students' *breadth of prior knowledge* about the topic was measured in keywords. Students had to write down as much keywords they could come up with concerning global warning in 15 minutes. The number of keywords was calculated and to determine the *quality of the prior knowledge* we compared the keywords they used to describe their knowledge of global warming with the keywords as agreed upon by the experts in the aforementioned Delphi procedure. To take into account spelling errors and minor word variations, we calculated a Levenshtein distance with a maximum allowed edit distance of two to determine a match between the expressed keywords and experts' set of keywords. This variable is called *expertise match score*.

Experienced task difficulty. The NASA Task Load Index (NASA TLX) was used to measure experienced task difficulty (Hart & Staveland, 1988). While originally developed for measuring the workload of operators, the NASA TLX is nowadays used in a wide variety of research domains (Hart, 2006) including multimedia research (Hilbert & Renkl, 2009; Wiebe, Roberts & Behrend, 2010; Zumbach & Mohraz, 2008). The NASA TLX consists of six subscales each containing a single self-report question: mental demand, physical demand, temporal demand, performance, effort, and frustration. The NASA TLX was slightly adapted by omitting the physical demand subscale, as the task did not require physical effort (cf. Zumbach & Mohraz, 2008). An open-ended question was added to the post-test questionnaire asking what specifically the participants found difficult in the task. Cronbach's alpha of the adapted NASA TLX was .65. The scale reliability increased to .71 after the performance subscale of the NASA TLX was excluded.

Tagging behavior. Tagging behavior was measured with three variables: 1) number of tags per bookmark, 2) unique number of used tags, and 3) the similarity of the used tags in the bookmark tool with the keywords mentioned in the prior knowledge measure.

The *number of tags per bookmark* was calculated as the average number of tags per bookmark for each participant. The *number of unique used tags* expressed the breadth of the used classification scheme for each participant. The similarity of the used tags in the bookmark tool with the prior knowledge of the participants was expressed in the *similarity prior knowledge – tag use score*. To calculate this score, we matched the used keywords from the prior knowledge test with the used tags when bookmarking webpages. Thus, this variable expresses the number of keywords from the prior knowledge of the participants that were also used as a classification. To take into account spelling errors and minor word variations, we calculated a Levenshtein distance with a maximum allowed edit distance of two to determine a match between a concept of the existing cognitive schema and the used tags for the bookmarks.

Use of bookmark tool. We used several variables to measure the use of the tagging tool. From the logging data we calculated the following time-logged metrics: 1) time spent on the pop-up and 2) time to first click on suggested tags. To analyze the eye movements, we defined four so called *areas of interest* (AOIs) for certain elements of the tool: 1) the page title, 2) the page url, 3) the used tags, and 4) the suggested tags. This is visualized in Figure 4. From the resulting eye-tracking data we extracted the following metrics: 1) the number of revisits for each AOI and 2) the fixation duration on each AOI.

Figure 4: AOIs on bookmark pop-up

The *eye tracking parameters* were calculated for all visits of the tool. Before further analyses, we filtered and processed the raw data with the SMI BeGaze software. We used the processed data to analyze the eye movements.

Procedure

Prior to the experiment, both, the participating students and their parents, received an introduction to the study. One month before the actual experiment the participants completed the paper and pencil pre-test questionnaire. The main experiment was conducted in individual sessions during the students' regular lessons. Each session started with a detailed explanation of the procedure. Next, the participants received an introduction about the eye tracking equipment. The eye tracker was adjusted to the individual features of each participant by calibrating the system with a 9-point calibration. To ensure the best possible data quality was obtained, we followed several guidelines (cf.: Holmqvist et al., 2011; Nyström et al., 2013): If participants wore either contact lenses or glasses and could easily read from a screen without them, we asked them to take part in the experiment without wearing the visual aids. If they could not easily read from the screen without visual aids, we asked them to wear contact lenses instead of glasses (if possible). Furthermore, we asked them to remove eye make-up, ty long hair at the back of their heads and sit as still as possible during the recording. Because recall accuracy of experienced workload decreases over time (Eggemeier & Wilson, 1991), the post-test questionnaire was administered directly after the end of the experiment with an online form. The duration of each session was approximately 50 minutes.

RESULTS

The results of this study are presented in three parts: existing knowledge schemata, experienced task difficulty, tagging behavior and the visual processing of the tagging tool. A t-test was used to analyze the differences between the two conditions for the measurements. A linear regression was used to analyze the effect of the expertise match score on the measurements as well as the effect of the interaction between the expertise match score and the condition on the measurements. Relevant means and standard deviations are summarized in Table 2.

Table 2 Means and standard deviations

	Given scheme	No scheme	p-value
N	34	34	
Prior knowledge			
Number of keywords	9.12 (3,92)	7.21 (2.67)	0.02
Expertise match score	3.44 (2.69)	2.76 (1.63)	> 0.05
Number of tags per bookmark	3.60 (2.13)	2.18 (2.07)	< 0.001
Number of unique used tags	15.62 (2.72)	20.03 (10.54)	0.02
Similarity prior knowledge - tag use	2.00 (1.28)	0.94 (1.35)	< 0.001
Use of bookmark tool			
Time on pop-up (seconds)	12.04 (10.47)	12.83 (10.87)	< 0.001
Time to first click Tags AOI (milliseconds)	3570.09 (4593.09)	1767.15 (2534.92)	< 0.001
Time to first click Suggestions AOI (milliseconds)	2677.56 (2478.31)	9483.45 (8942.15)	< 0.001
Fixation duration Title AOI (milliseconds)	117.42 (269.30)	370.37 (878.94)	< 0.001
Fixation duration URL AOI (milliseconds)	51.08 (183.55)	143.54 (517.26)	< 0.001
Fixation duration Tags AOI (milliseconds)	598.52 (905.08)	2595.42 (3094.42)	< 0.001
Fixation duration Suggestions AOI (milliseconds)	6042.04 (4752.30)	3631.74 (4467.12)	< 0.001
NASATLX			
Mental demand	3.68 (1.22)	3.82 (1.14)	> 0.05
Temporal demand	4.21 (1.53)	4.74 (1.26)	> 0.05
Performance	5.29 (1.12)	5.00 (0.98)	> 0.05
Effort	3.71 (1.31)	4.03 (0.97)	> 0.05
Frustration	2.24 (1.21)	2.41 (1.35)	> 0.05
Total NASA TLX score	19.12 (4.10)	20.00 (3.27)	> 0.05

Differences in existing schemata and their use

From the prior knowledge measures we derived the keywords participants used to express their existing knowledge about the topic of global warming and matched these with the keywords as agreed upon by the experts. This resulted in an expertise match score. The quality of the prior knowledge is an important determinant for the performance of the participants; this measure is a key input variable to evaluate the stated hypotheses. Although the given classification scheme group (M = 3.44, SE = 0.46) had a higher score than the no- classification scheme group (M = 2.76, SE = 0.28), the difference between the two groups was not significant ($t(54.46) = -1.25$, $p > .05$). The number of used keywords correlates significantly with the expertise match score, $r = .48$, $p < .001$. This is not surprising as a greater number of keywords increases the change of a match with the formulated expert schema.

Experienced task difficulty

The experience task difficulty was measured with five subscales of the NASA TLX instrument. For all subscales except one, participants in the given classification group had a lower score. However, the two groups did not differ significantly on the subscales of mental demand ($t(65.68) = 0.51$, $p > .05$), temporal demand ($t(63.66) = 1.55$, $p > .05$), performance ($t(65.00) = -1.15$, $p > .05$), effort ($t(60.67) = 1.15$, $p > .05$), and frustration ($t(65.19) = 0.57$, $p > .05$) or on the total task load index ($t(62.89) = 0.98$, $p > .05$). Additionally, we analyzed the experienced task difficulty with the expertise match score as predictor variable. For the subscales mental demand, temporal demand, performance and effort, these analyses did not result in significant differences. Participants with a higher expertise match score experienced a higher level of frustration ($R^2 = .07$, $\Delta R^2 = .06$, $F(1, 66) = 5.075$, $p = .028$), but adding condition as a predictor did improve the prediction ($R^2 = .10$, $\Delta R^2 = .07$, $F(2, 65) = 3.526$, $p = .035$) without having a significant effect of condition ($p > .05$).

Tagging behavior

As described before we measured tagging behavior with three variables: the number of tags per bookmark, the number of unique used tags, and a similarity prior knowledge – tag use score. We analyzed these variables taken into account differences between conditions, the number of keywords and expertise match score.

With regard to *number of used tags per bookmark*, participants in the given schema group used significantly more tags for each bookmark ($t(1231.3) = -11.89$, $p < .001$, $r = .33$) than participants in the no-classification scheme condition. This difference had a medium sized effect. The expertise match score by itself had not a significant influence on the number of used tags per bookmark ($R^2 = .002$, $\Delta R^2 = .001$, $F(1, 1240) = 2.247$, $p > .05$). Adding the condition as predictor increased the predictive value of the model considerably (R^2 = .11, ΔR^2 = .11, F(2, 1239) = 76.22, p < .001) with both predictors having a significant effect. However, for the expertise match score predictor this effect remained fairly low as could have been expected after the insignificant effect when used as the only predictor.

Although participants in the given classification scheme condition used more tags for each bookmark, they used significantly less *unique tags* in the whole task compared to the no-classification scheme condition (t(37.37) = 2.36, p = .02, r = .36). No evidence was found for a linear association between the expertise match score and the number of unique tags (R^2 = .001, ΔR^2 = -0.01, F(1, 66) = .06, p > .05). Adding condition did not result in a model that explained the number of unique tags better (R^2 = .08, ΔR^2 = .05, F(2, 65) = 2.75, p > .05).

The degree to which participants used *tags similar to the keywords from prior knowledge* was significantly higher for the given classification scheme condition in comparison with the no-classification scheme condition (t(65.82) = -3.32, p = .001, r = .38). The expertise match score had a significant positive relation with number of tags used (R^2 = .13, ΔR^2 = .12, F(1, 66) = 10.07, p = .002). Following the addition of condition resulted in a significant increase (F(1, 66) = 9.13, p = .004) of the predictive value of the model (R^2 = .24, ΔR^2 = .22, F(2, 65) = 10.22, p < .001). The effect of condition in the latter model was much larger than the effect of the expertise match score.

Use of the bookmark tool

The use of the bookmark tool is described with several variables. In the first part of this paragraph we report three time-logged measurements: time spent on the pop-up, time to first click on the tags AOI and time to first click on the suggestions AOI. Next, we report viewing behavior on the tagging tool for the page title AOI, the url AOI, the tags AOI and the suggestions AOI. For each of these measurements the differences between the two conditions and the quality of prior knowledge measured by the expertise match score will be taken into consideration.

The *time spent on the pop-up* was significantly longer for participants in the in the no-classification scheme condition than participants in the given classification scheme condition (t(8299.2) = 3.41, p < .001, r = .04). However, the effect size was rather small. Using the expertise match score as a predictor resulted in a significant, but small effect (R^2 = .00, ΔR^2 = .00, F(1, 7772) = 4.717, p = .03). Including condition improved the model (R^2 = .003, ΔR^2 = .002, F(2, 7771) = 10.07, p < .001) significantly (F(1, 7771) = 15.42, p < .001). However, the explanatory power of both regression models remained very low.

The *average time the participants needed for the first click inside the tags AOI* was considerably and significantly longer for participants in the given classification scheme condition (t(95.43) = 3.41, p < .001, r = .34). Regression models with only the expertise match score (R^2 = .00, ΔR^2 = .00, F(1, 529) = 0.00, p > .05) or with condition include (R^2 =

.03, ΔR^2 = .03, $F(2, 528) = 8.486$, $p < .001$) showed no or only a small effect on the dependent variable, although the latter model was an improvement over the model with only the expertise match score as a predictor ($F(1, 528) = 16.97$, $p < .001$).

The *average time the participants needed for the first click inside the suggestions AOI* was considerably and significantly longer for participants in the no classification scheme condition ($t(669.11) = 18.12$, $p < .001$, $r = .57$). A regression model with only the expertise match score showed a very small but significant effect ($R^2 = .00$, $\Delta R^2 = .00$, $F(1, 1274) = 5.404$, $p = .02$). Including also condition showed a considerable improvement ($F(1, 1274) = 360.12$, $p < .001$) of the model ($R^2 = .22$, $\Delta R^2 = .22$, $F(2, 1273) = 183.5$, $p < .001$). However, the effect of the variable expertise match score was not significant.

Participants in the no-classification scheme condition spent on average more time on the title AOI than participants in given classification scheme condition ($t(829.29) = 7.43$, $p < .001$, $r = .25$). The expertise match score had no significant effect on the time spent on the title AOI ($R^2 = .00$, $\Delta R^2 = .00$, $F(1, 1440) = 0.259$, $p > .05$). Adding condition resulted in a better model ($R^2 = .04$, $\Delta R^2 = .04$, $F(2, 1439) = 32$, $p < .001$), but only the condition was significant.

Participants in the no-classification scheme condition spent on *average more time on the url AOI* than participants in given classification scheme condition ($t(867.34) = 4.56$, $p < .001$, $r = .15$). The expertise match score had no significant effect on the time spent on the url AOI ($R^2 = .00$, $\Delta R^2 = .00$, $F(1, 1440) = 0.306$, $p > .05$). Adding condition resulted in a considerable improvement ($F(1, 1440) = 27.31$, $p < .001$) of the model ($R^2 = .02$, $\Delta R^2 = .02$, $F(2, 1439) = 32$, $p < .001$), but only the condition was significant.

Participants in the no-classification scheme condition spent on *average more time on the tags AOI* than participants in given classification scheme condition ($t(819.36) = 16.71$, $p < .001$, $r = .50$). The expertise match score had no significant effect on the time spent on the tags AOI ($R^2 = .00$, $\Delta R^2 = .00$, $F(1, 1440) = 0.501$, $p > .05$). Adding condition resulted in a considerable improvement ($F(1, 1440) = 297.32$, $p < .001$) of the model ($R^2 = .17$, $\Delta R^2 = .17$, $F(2, 1439) = 149$, $p < .001$), but only condition was significant.

Participants in the no-classification scheme condition spent on *average much less time on the suggestions AOI* than participants in given classification scheme condition ($t(1555.1) = -10.37$, $p < .001$, $r = .25$). The expertise match score had no significant effect on the time spend on the tags AOI ($R^2 = .00$, $\Delta R^2 = .00$, $F(1, 1440) = 1.176$, $p > .05$). Adding condition resulted in a considerable improvement ($F(1, 1440) = 89.26$, $p < .001$) of the model ($R^2 = .06$, $\Delta R^2 = .06$, $F(2, 1439) = 45.25$, $p < .001$), but only condition was significant.

DISCUSSION

The aim of this study was to get more insight into whether providing support with an expert defined classification scheme for tags had an effect on storing and organizing information found on the Internet with bookmarks. For this study, participants received a simulated task in which they collected information to prepare a fictional newspaper article on global warming for a large national newspaper. Participants stored and organized the found information with a bookmarking tool. Besides providing a tag classification support or not, prior knowledge on the topic was an important factor. Results showed that there were differences, although not significant, in prior knowledge between the experimental and control group. Analysis of the process of storing and organizing information consisted of three parts: the effects on experienced task difficulty, tagging behavior, and the use of the bookmarking tool.

Based on previous research we expected that higher prior knowledge would lead to lower experienced task difficulty (H1a) and lower prior knowledge would result in lower experienced task difficulty when receiving support in the given classification condition and higher experienced task difficulty in the no-classification condition (H1b). These expectations were not completely supported by the results. Although the experienced task difficulty was lower for the given classification group, the differences between the no-classification group and the given classification group were not significant. As it takes time and effort to develop expertise on a certain topic (e.g., Clark et al., 2012; Kalyuga, 2007; Kalyuga et al., 2003; Kirschner et al., 2006; Van Merriënboer & Kirschner, 2012), variation in prior knowledge among participants might not have been pronounced enough.

With regard to tagging behavior we expected an interaction of the level of prior knowledge and condition on the number of used (unique) tags. We hypothesized that participants with higher prior knowledge would use more tags per bookmark (H2a) and more unique tags during the task (H2b). Participants with lower prior knowledge would use (relatively) more tags in the given classification condition and less tags in the no classification condition (H2c). Low prior knowledge participants were expected to use less unique tags in both conditions (H2d). Furthermore, participants in the given classification condition were expected to use more tags which are similar to their prior knowledge (H2e). Contrary to our expectations, we did not see a relation between prior knowledge and these measures. With regard to differences between the two conditions, results were mixed in relation to the hypotheses. Participants in the given classification condition used more tags per bookmark, which is in line with the expectation (H2c). When you do not have much prior knowledge on the topic, it is much harder to classify a new information source because you do not have an existing cognitive schema to which the source can be related to (Clark et al., 2012; Rogers & Swan, 2004; Walhout, et al., submitted). Thus, providing support in the form of a classification scheme, can help the participants to elicit tacit knowledge because they recognize the used words in the classification scheme which they could not have come up with themselves (Clark et al., 2012). Participants in the given classification condition used less unique tags during the task. This finding was unanticipated. A possible explanation for this might be that providing support to novices gives them direction and thus focus, while at the same time novices without support have no clue about what to do (Clark et al., 2012). This is also corroborated by Chen, Chen and Sun (2014) who found that including a tag-based support system improved reading efficiency and comprehension. These participants likely made the mistake of creating too specific tags that are only useful for one or a few sources (Guy & Tonkin, 2006). As a result of making such a mistake, participants end up with a large set of unique tags that each have limited use. For the third measurement of tagging behavior, participants in the given classification condition used more tags that were similar to keywords from their prior knowledge statements compared to the participants in the no-classification condition. This is consistent with what was expected (H2e). Confronting participants with a classification in the given classification condition allowed them to recognize familiar concepts and connect them to the information source they wanted to store.

Having discussed the actual tagging behavior of the participant, in this section we will address the formulated hypotheses about how the participants processed the bookmarking pop-up. We expected an interaction between prior knowledge and the experimental condition. High prior knowledge would lead to less time processing the bookmarking tool (H3a) and less time for the first interaction with the tool (H3b) in the given classification condition. Low prior knowledge would in our expectations result in more time on the tool (H3c) and more time for the first interaction with the tool (H3d). Again, we found no evidence for an influence of prior knowledge, either by itself or in conjunction with the experimental condition, on the way the participants processed the bookmarking pop-up. However, we found several significant differences between the two groups with regard to the time logged measures. Although participants in the given classification group needed less time to process the pop-up, they spend significantly more time evaluating the suggested tags given in the pop-up while at the same time only looking shortly at the tags they selected. These results are a strong indication that the provided support is actually used by the participants. These results also suggest that participants in the no-classification condition also considered the tags they previously used when storing a new bookmark. However, they still spent a considerable amount of time thinking about how to classify the bookmarks. Participants in the no-classification group also looked significantly more at other parts of the pop-up, such as the title and the url of the webpage or the whitespace outside the designated AOIs. The pop-up contained two clues about the source contents: the title of the page and the url. Besides spending a considerable amount of time looking at the field in which they had to enter the tags, they also looked more at the title and url fields. Furthermore, participants spent nearly twice as much time outside the predefined AOIs. All these differences in viewing behavior show that giving support inside the bookmarking pop-up leads to a different way of processing of the bookmarking pop-up.

The results of this study indicate that giving students support in classifying information sources influences their behavior and also possibly improves their information storing and organizing behavior. Supporting students with a classification scheme might therefore benefit students in solving information problems for educational assignments. Ultimately, this might lead to a better development of cognitive schemata. This will especially be true for novices. Ideally, these supporting classifications should be developed in cooperation with experts who formulate the keywords as much as possible in laymen's terms. Students can then easier relate to the expert-defined support. When students develop more knowledge about a subject, such support should probably be reduced (Van Merriënboer & Kirschner, 2012).

CONCLUDING REMARKS

Searching the Internet is a complex process, which receives continuous attention in academic research (Brand-Gruwel, et al., 2009). As mentioned in the introduction, the subprocess of storing and organizing information during the search process, however, has not received much attention. In this study, we investigated how providing support influences the processes of storing and organizing found information. The main conclusion is that providing support has a positive influence on the number of tags used to store a bookmark and that giving support elicits the use of concepts which were already part of the prior knowledge of the participants. Furthermore, the use of less unique tags indicates a more focused use of tags to organize the information. This might indicate that providing support leads to a less cluttered classification of the collected information. However, to give a more conclusive answer, more research and other types of analysis are needed.

In this study, we found no evidence for the influence of prior knowledge. A possible reason for this might be the level of expertise of the participants. It takes time and effort to develop expertise on a certain topic (e.g.: Clark et al., 2012; Kalyuga, 2007; Kalyuga et al., 2003; Kirschner et al., 2006; Van Merriënboer & Kirschner, 2012). Although the prior knowledge of participants varied considerably, it might be that the used sample of participants is still relatively homogeneous because they were only recruited from the highest level of secondary education in The Netherlands.

Although we addressed some aspects of storing and organizing information and the influence of providing support, several questions remain to be answered. Even though we found no evidence for the influence of prior knowledge, this should be studied in more depth. Including participants from other levels of secondary education could result in a more diverse sample of participant with regard to prior knowledge on the subject of global warming. However, even that might not be enough to make a clear distinction with regard to expertise level. Consequently, other types of participants could be included as well. Examples are including other age groups in the experiment or first year students from university and experts. Another area on which future research could elaborate is the used tags in both conditions. A refinement of the given classification condition could be to allow participants to add their own tags. It would be interesting to see how many tags are added on top of the provided classification schema. Furthermore, a content analysis of the created tags (also in the no-classification condition) could give more insight in the tagging behavior of different groups of users. Finally, it could be useful to see whether the first use effect as demonstrated by Golder and Huberman (2006) and Alawami (2014) also holds for given classifications. Besides further research on the issue of how to organize found information, also the effect of providing a classification on *search behavior* could be an interesting path. We would expect students to use words from the classification scheme in their search queries.

The implications of the results of this study for educational practice are that when designing IPS assignments for students, especially novices, also to give attention to the supporting information that comes with the assignment. A good guideline for designing such assignments could be the 4CID model (Van Merriënboer & Kirschner, 2012). In this design approach the diminishing role of supporting information matches the increase of knowledge of the students.

Chapter 6

General discussion

Nowadays, the Internet is used by many as the main source of information (e.g., Jansen & Spink, 2006; Rouet et al, 2011). Although students regularly use the Internet to find an answer to their problem, the process of finding that answer is not without problems (e.g., Bilal, 2000; Brand-Gruwel et al., 2005; MaKinster et al., 2002; Large & Beheshti, 2000; Van Strien et al., 2014; Walraven et al., 2008). Students often do not know which search terms to use when searching the Internet, how to judge the validity of websites, or how to question the source. Instead, the choice for opening a site is mainly guided by the title or summary of the site. Mastering this is part of so-called Information Problem Solving (IPS). There is little doubt that IPS skills are essential in the information abundant society we are living in. It is therefore often argued that teaching IPS skills should be part of modern education (e.g., Brand-Gruwel et al., 2009; Van Deursen & Van Dijk, 2008). Over the past decades, a substantial body of research on IPS has emerged. Based on this research, several prominent models like the search process model (Kuhlthau, 2004), the Big6-model (Eisenberg & Berkowitz, 1990) and the IPS-I model (Brand-Gruwel et al., 2009) have been developed.

In addition to deliberately searching for information, we also often accidentally stumble upon new information on the Internet. Whether it is a result of a deliberate search process or an unintended encounter with new information, we have to decide what to do with the information we deem to be useful and we want to re-access at a later time (Whittaker, 2011). The question of re-access is also related to the purpose of the search for information. A simple navigational search is less likely to evoke a decision about re-access than a search for information when solving a more complex problem where information has to be compared with previously found information or when information is collected over time (Rouet, 2003). However, in contemporary research into IPS skills, the aspect of storing and organizing information has received little attention. The main goal of this thesis was to fill this research gap. Based on the IPS-I model (Brand-Gruwel et al., 2009), the studies presented in this thesis investigate in depth the process of storing and organizing information while searching on the Internet.

MAIN FINDINGS

The study presented in Chapter 2 focused on the relation between task complexity and search behavior. Evaluation of information during information problem solving processes already starts when trying to select the appropriate search result on a search engine results page (SERP). In this study we investigated the influence of task complexity on search query formulation, evaluation of search results, and task performance by means of a within-subjects design. Participants performed three search tasks, namely a fact-finding, cause–effect, and a controversial topic task. Mosenthal's framework of prose task characteristics served to assess each task's complexity (Mosenthal, 1998). This framework is composed of three different dimensions: type of information requested, type of match, and plausibility of distractors. During the actual experiment, participants could search the Internet without any restrictions or time limit. Their perceptual search processes were measured with a combination of log files, eye-tracking data, answer forms, and think-aloud protocols. Results revealed that an increase in task complexity led to an increase in interaction with the search engine: more search queries and used keywords, more time to formulate search queries, and more considered search results on the SERPs. In addition, task performance decreased when task complexity increased. Furthermore, higher ranked search results were considered more often than lower ranked results. However, not all the results for the most complex task were in line with expectations. For the most complex task, the interaction dropped as well, while the time on task remained the same as in the simpler tasks. These results can be explained by a lack of prior knowledge and the possible interference of prior attitudes. This leads to the conclusion that search tasks should be designed such that the complexity of the task and the accompanying support are adapted to the prior knowledge and skill levels of the students (Van Merriënboer & Kirschner, 2012).

While the study in Chapter 2 looked at the search process when searching on the Internet, the study presented in Chapter 3 looks at searching inside a collection of information sources and how different ways to support navigating through such a collection influence performance and behavior. Relatively recent, tag clouds were introduced to navigate large collections of information. Up till then, navigational structures were mostly hierarchical in nature. Tag clouds are considered more flexible and provide a better description of the contents (Civan et al., 2009). Although previous research has compared several types of navigational support extensively, few studies exist that compare hierarchical structures with tag clouds directly (Trattner et al., 2012; Voit et al., 2012). In this study, we investigated the effect of a navigation tool (navigational support with either a tag cloud or conventional hierarchical menu), task complexity (fact-finding vs. information-gathering task) and the user's gender on navigation behavior, visual processing of the navigation tool and task performance. To study these effects, we constructed two almost identical learning environments. The only aspect in which these learning environments differed was the type of navigational support, which was either a hierarchical menu or a tag cloud. Participants were randomly assigned to one of these two conditions such that the gender balance was equal for both conditions. Results showed that neither the type of navigation nor gender could explain differences in task performance for both tasks. However, participants differed in their information processing. In the tag cloud condition, participants needed more time to process the navigational tool and less time to process the overview pages. Furthermore, they also revisited pages less often. The deeper processing of information needed for the information-gathering task was reflected in fewer visits to, but longer viewing times of pages. Although participants were not familiar with using tag clouds to navigate content, their performance did not suffer from this unfamiliarity. Tag clouds can therefore be viewed as a viable alternative to the more traditional hierarchical menus.

Chapter 3 gave us more insight in how students navigate through a collection information sources. However, compiling such a collection occurs during the search process where in the sub-process of storing and organizing information the collection is build. When encountering useful information, students are confronted with the question how to keep that information so they can use it at a later time. The study presented in Chapter 4 combined several aspects of the first two studies. As in the first study, the participants received a search task for which they collected information on a certain subject (obesity). They were asked to store the information which they judged as useful with a bookmarking tool in which they could use a hierarchical structure or a tag cloud to store the information (like in the second chapter). With this study we sought to gain more insight on how classification support (no vs. given classification) and support with an organizing system (hierarchical vs. tag cloud) influence bookmarking behavior. Additionally, we also looked at the interaction with prior knowledge. The results showed that neither classification support nor the type of organizing system resulted in selection of webpages of higher quality or differences in the diversity of the used classifications. Based on previous research (Kalyuga, 2007; Kalyuga et al., 2003; Rogers & Swan, 2004; Špiranec & Ivanjko, 2013; Stadtler & Bromme, 2008) we expected that participants who received classification support would perform better than participants who did not receive such support. Because participants have little experience with a tagging system to organize information, one could argue that participants in the hierarchical conditions had an advantage because they were able to use system they are used to. As also previous studies comparing hierarchical and tagging systems reported mixed results, no clear hypothesis about which one is better would be formulated.

Also in this study neither of the two organizing systems led to better results. That means that a tagging system is a viable alternative for the use of a hierarchical system to organize information. Although most people are unfamiliar with using a tagging system, once used, they indicate that they like the greater flexibility, the better descriptive characteristics and easier connection with related concepts (Bergman et al. 2013b; Civan et al., 2008; Voit et al., 2012). Chapter 5 presents a follow-up study of Chapter 4. In this study, we choose to focus on only the tagging system. This would allow us to better isolate the effect of support with a classification scheme. The aim of this study was to determine the extent in which providing a classification scheme has an effect on experienced task difficulty, tagging behavior and how participants used the given bookmark tool. Furthermore, we also considered the interaction with the level of prior knowledge. The main conclusion of this study is, that providing students support with a classification scheme to organize information influences their behavior and also possibly improves their information processing. Providing support had a positive influence on the number of tags used to store a bookmark and improved the use of concepts which were already part of the participants' prior knowledge. Furthermore, the use of less unique tags indicates a more focused use of tags to organize the information. This might indicate that providing support leads to a less cluttered classification of the collected information.

Providing support will especially benefit novices (e.g., Clark et al., 2012; Kalyuga et al., 2003; Kalyuga, 2007; Kirschner et al., 2006; Van Merriënboer & Kirschner, 2012). Such a classification support should be preferably designed in cooperation with topic experts. The supporting classification scheme should be formulated as much as possible in lay-men's term, thus enabling students to easier relate to the expert defined support.

DISCUSSION AND IMPLICATIONS

This thesis explores some ways in which people, and students in particular, can be supported in storing and organizing interesting information from the Internet. The key components are whether it is better to support people with a hierarchical or a tagging system and whether supporting them with a pre-determined classification scheme is beneficial. The presented research in this dissertation addresses four main topics with regard to the search phase of the IPS process. Specifically, the several studies focus on the sub processes of storing and organizing information. The four main topics are (1) the influence of task complexity, (2) the influence of type of organizer, (3) the influence of providing support with a classification schema, and (4) the interaction with prior knowledge. In the following sections each topic will be discussed.

Influence of task complexity

Task complexity was the main topic of interest in Chapter 1 and one of the variables under consideration in Chapter 2. From previous research we know that an increase in task complexity generally leads to a decrease in task performance (e.g., Byström & Järvelin, 1999; Clark et al., 2012; Jansen & Spink, 2006; Singer et al., 2012a; Vakkari, 1999). Both studies found that participants provided less correct answers for the more difficult tasks and thus confirmed the findings of a large body of previous research. However, task performance can not only be expressed by whether or not the correct answer was given, but also by other measures like time on task (Borlund & Dreier, 2014; Brennan, Kelly & Arguello, 2014; Jansen & Spink, 2006; Kelly et al., 2015; Liu et al., 2010) or specific time-based measurements (cf. Juvina & Van Oostendorp, 2006; Singer et al., 2012a) can be used to further assess the performance of the participants. In both studies of the current thesis, an increase in task complexity from a simple fact-finding task to a more complex search task led to an increase in time on task. Also, both studies showed a decrease in browsing activity in relation to time on task with increasing task complexity. However, as the first study showed, the relationship between task complexity and several measurements is not linear. Based on an assessment with Mosenthal's framework of prose task characteristics (Mosenthal, 1998), the last task in that study was more difficult than the second task. Nonetheless, this did not lead to a further increase in time on task. That third task was probably too difficult for the participants and they probably gave up on trying to find a solution. Indications for that are that they formulated much less search queries and evaluated the SERPs not as extensively as in the second task. In addition to being too difficult, prior attitude might have had a mediating effect as well. Past research has shown that prior attitudes have an influence on the IPS-process (e.g., Brannon, Tagler & Eagly, 2007; Fischer & Greitemeyer, 2010; Van Strien, Brand-Gruwel, & Boshuizen, 2014; White, 2014). While these two studies confirmed the relationship between task complexity and task performance, they also showed that there are some nuances in this relationship.

Influence of type of organizer

The type of organizing system was the key concept under consideration in Chapters 3 and 4. As stated in both chapters, traditionally, information was organized hierarchically, which has several limitations with regard to flexibility and re-findability (Bush, 1945; Wright, 2007). In contrast to the hierarchical systems we are used to, a tagging approach is flat and allows for multiple classifications (Shirky, 2005; Wichowski, 2009). The studies in Chapters 3 and 4 compared the hierarchical and the tagging approaches in two aspects of information problem solving: re-accessing information (Chapter 3) and storing and organizing information (Chapter 4). Both studies showed that neither of the two systems of organizing information could be considered as better. A disadvantage of the tagging system was that participants were hardly familiar with it. Although people express a preference for tags when explicitly asked (Bergman et al, 2013b; Civan et al, 2009), they fall back to using a hierarchical system in their actual behavior. Learning something new requires effort, thus they rely on the system they are used to. Nonetheless, participants in both studies who did use a tagging system, did not perform worse compared to participants using a hierarchical system. Although some researchers assume that familiarity is a key aspect in inhibiting people to start using a tagging system (Bergman et al, 2013a & 2013b), this was not explicitly investigated. Despite the fact that they did not make a comparison with a hierarchical system, Lee at al. (2009) found that a higher familiarity with a tagging system leads to better performance. It can therefore be hypothesized that experience with tagging might change the conclusions drawn thus far. Additionally, the way in which tags are used might also influence the outcome. As shown by Basile et al. (2014) and Fastrez and Jacques (2015), using tags to describe the contents of an information source is probably the best fit for a tagging system.

Influence of using a classification scheme

Having prior knowledge about a topic is advantageous in solving related information problems (Clark et al., 2012; White et al., 2009). Previous research has shown that experts and novices differ considerable in behavior and performance when solving information problems (e.g., Brand-Gruwel et al., 2017; Chen & Macredie, 2010; Monchaux et al., 2015; Lee & Pang, 2017; White et al., 2009). The prior knowledge of experts is structured in well-developed cognitive schemata (Bergman, et al., 2013; Kirschner et al., 2006), which are used to compare new information with what is already know (Kirschner et al., 2006; Rogers & Swan, 2004). Consequently, experts can process new information much easier than novices as the latter do not have such well-developed internal knowledge structures. Therefore, supporting novices with a classification scheme might improve their performance (De Vries et al., 2008; Stadtler & Bromme, 2008; Van Merriënboer & Kirschner, 2012). The influence of providing novices with a classification scheme was subject of the studies presented in Chapters 4 and 5. Both studies looked at behavioral effects. One of the most profound effects in both studies was, that participant receiving support with a classification scheme also used more tags which were already present in their prior knowledge structures. Thus, the classification scheme allowed participants to recognize familiar concepts and relate those to the information they wanted to store and organize. Furthermore, the support with a classification scheme led to using classifications more often and a more focused use of classification keywords (Chapter 5).

Interaction of prior knowledge with type of organizer and using a classification scheme

In the studies presented in Chapters 4 and 5 we included measurements of prior knowledge to study whether this had an effect on performance. In both studies, we found no evidence of an influence of prior knowledge on performance or on behavioral measures. An explanation for this could be that participants in these two studies were neither experts nor even intermediates. Although the level of prior knowledge varied considerably among participants, this variation is still quite low when compared to experts. To become an expert on a certain subject requires a lot of time and effort (e.g., Clark et al., 2012; Kalyuga et al., 2003; Kalyuga, 2007; Kirschner et al., 2006; Van Merriënboer & Kirschner, 2012). The participants in our studies were students from secondary education and can therefore be considered novices on the subjects used in these experiments. To better assess the effect of prior knowledge, future experiments should compare more heterogeneous groups of participants.

IMPLICATIONS FOR FUTURE RESEARCH AND CONCLUSIONS

In each of the chapters in this dissertation, specific limitations and suggestions for further research were discussed. In this section, these will be synthesized for this thesis as a whole. Because limitations are also implicit suggestions for future research, the limitations and suggestions will be discussed in integration.

As highlighted in the last part of the previous paragraph, future research might specifically be designed to account for a greater variation in prior knowledge. For all the presented studies, students from the highest level of secondary education were selected. Although participants varied with regard to their level of prior knowledge, this variation was still rather homogeneous. Future research could achieve a greater variation in several ways. First, the easiest way to get a greater variation in prior knowledge is to include students from other levels of secondary education. The advantage of this approach is that it is easy to include participants of the same age group. Second, as building up expertise needs a considerable investment in time and effort (e.g., Clark et al., 2012; Kalyuga et al., 2003; Kalyuga, 2007; Kirschner et al., 2006; Van Merriënboer & Kirschner, 2012), it can be worthwhile to included intermediates and true experts on a topic to distinguish effects of prior knowledge. Example of this approach within the field of IPS research are the studies of Brand-Gruwel et al. (2005) where first year students from the university were compared with PhD students and Brand-Gruwel et al. (2017) where a group of first year psychology students were compared with a group of assistant professors. Other examples are the studies by Duggan and Payne (2008) and Sanchiz, Chevalier and Amadieu (2017), who also showed that domain knowledge helps in finding information on the Internet.

In this thesis two studies directly compared the hierarchical system with the tagging system. Neither of these studies which compared hierarchical and tagging systems, resulted in conclusive results with regard to whether one or the other should be preferred. Because research on this comparison is quite scarce, more research is needed to get a better insight in the differences between those two approaches of organizing. Future research should especially take into account the low familiarity of people with the tagging system. As discussed earlier, the low familiarity of participants with tagging approaches requires some kind of familiarization period before people take part in an experiment which compares the tagging system to the hierarchical system. One of the few examples of this is the study by Bergman et al (2013a) which used a short familiarization period. Although from a practical viewpoint the familiarization period cannot be too long, from an experimental design viewpoint it should be considerably longer than Bergman et al (2013a) used to enable a fair comparison.

Today, a substantial body of research exists into design choices for hierarchical menus (e.g., Bezdan et al., 2013; Leuthold et al., 2011; Puerta Melguizo et al., 2012). On the other hand, few studies exist that investigated how to construct and visualize tag clouds (Trattner, et al., 2012). Studying and comparing variations in visual presentation and the level of detail will provide more insight in whether certain design decisions affect behavior and performance positively or negatively.

Providing participants support with a classification scheme is beneficial with regard to organizing the found information. However, task performance was not included in the measurements. The change in behavior due to the classification support is a first step, but ultimately we want to know whether support with a classification scheme also leads to better performance. Another effect of providing support with a classification scheme might be the interaction with the search process. Although outside the scope of the research for this thesis, it is an interesting aspect to investigate. One can easily hypothesize that giving students support with a classification scheme, might also influence the keywords they use for subsequent search queries.

In three of the four studies in this thesis, visual processing was studied with the use of eye-tracking. Differences in viewing behavior do not necessarily have to mean that there are also differences in cognitive processing. An interesting study in this regard is the research by Bhattacharya and Gwidzka (2018) who showed that differences in learning in online search tasks were related to differences in viewing behavior. It would be interesting to see how support with a classification schema interacts with this.

Searching for information is increasingly shifting to mobile devices (Sterling, 2015). This development gives people even more the impression that all information is available at their fingertips. However, search behavior on mobile devices is inherently different from search behavior on PC's (Song et al., 2013), thus making mobile devices important for research in IPS. Whether a hierarchical or a tagging system should be preferred on mobile devices is uncertain and consequently needs to be studied.

Finally, the affluence of information on the Internet has also enabled a widespread creation of conspiracy theories (Del Vicario et al., 2016; Lewandowsky, Oberauer & Gignac, 2013; Sunstein & Vermeule, 2009). An important cause of the creation of conspiracy theories is lack of information and topic knowledge (Sunstein & Vermeule, 2009). Supporting students with expert defined classification schemata will help them to construct their own cognitive schemata more accurately and faster. A better construction of cognitive schemata could help students better judge the information around them. Consequently, they might be less susceptible to conspiracy theories. However, studying this effect is not an easy endeavor as it requires working with a large group of participants over a long period of time.

Whether we search for the longest river in Africa or gather information on how to deal with a disease, the Internet is the source to find any piece of information. Being able to search, find, evaluate, select, process, organize and present information to solve a problem are defined as information problem solving skills (e.g., Brand-Gruwel et al., 2009). The research presented in this thesis contributes to the still growing body of research into this topic. We learned about the search process that the relation between task complexity and performance is not linear. When tasks become too difficult, the dynamics in behavior change. About storing and organizing information, the main topic of this thesis, we can conclude that using tags are a viable alternative to hierarchical approaches. We can even assume that when people are more familiar with tagging, their performance will improve. Studies which incorporate a substantial familiarization period should be done to test this assumption.

References

Abrams, D., Baecker, R., & Chignell, M. (1998). Information archiving with bookmarks: personal Web space construction and organization. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 41-48). ACM. New York, NY, USA.

Adar, E., Teevan, J., Dumais, S. T., & Elsas, J. L. (2009). The web changes everything: understanding the dynamics of web content. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining* (pp. 282-291). ACM. New York, NY, USA.

Ahmed, I., & Blustein, J. (2006). Influence of spatial ability in navigation: using look-ahead breadcrumbs on The Web. *International Journal of Web Based Communities, 2*, 183-196. doi: 10.1504/IJWBC.2006.010309

Alawami, A. (2014). Bookmarking and Tagging Patterns in Social Bookmarking Systems. In *Proceedings of the International Conference on Information and Knowledge Engineering (IKE)* (p. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help Seeking and Help Design in Interactive Learning Environments. *Review of Educational Research, 73*, 277–320. doi: 10.3102/00346543073003277

Alhenshiri, A. (2013). Studying the User Task of Information Gathering on the Web (Doctoral dissertation). Retrieved October 9, 2016, from: https://dalspace.library.dal.ca/bitstream/handle/10222/21421/Alhenshiri-Anwar-PhD-ComputerScience-March-2013.pdf.

Alhenshiri, A., Watters, C., Shepherd, M., & Duffy, J. (2012). Building support for web information gathering tasks. In System Science (HICSS), 2012 45th Hawaii International Conference on (pp. 1687-1696). IEEE.

Amadieu, F., Van Gog, T., Paas, F., Tricot, A., & Mariné, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learning and Instruction, 19*, 376-386. doi: 10.1016/j.learninstruc.2009.02.005

Andreano, J. M., & Cahill, L. (2009). Sex influences on the neurobiology of learning and memory. *Learning & Memory, 16*, 248-266. doi: 10.1101/lm.918309

Andreesen, M. (1993, Februari 16). New X-based information systems browser available. [Newsgroup]. Retrieved from: https://groups.google.com/forum/#!topic/comp.infosystems/855ZNG7mvI0 on September 13, 2015.

Anfinnsen, S., Ghinea, G., & De Cesare, S. (2011). Web 2.0 and folksonomies in a library context. *International Journal of Information Management, 31*(1), 63-70. doi: 10.1016/j.ijinfomgt.2010.05.006

Arolas, E. E., & Ladrón-de-Guevar, F. G. (2012). Uses of explicit and implicit tags in social bookmarking. *Journal of the American Society for Information Science and Technology, 63*(2), 313-322. doi: 10.1002/asi.21663

Aula, A., Jhaveri, N., & Käki, M. (2005). Information search and re-access strategies of experienced web users. In *Proceedings of the 14th international conference on World Wide Web* (pp. 583-592). ACM. New York, NY, USA.

Aula, A., & Nordhausen, K. (2006). Modeling successful performance in Web searching. *Journal of the American Society for Information Science and Technology*, *57*(12), 1678-1693. doi: 10.1002/asi.20340

Aula, A., Khan, R. M., & Guan, Z. (2010). How does search behavior change as search becomes more difficult?. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 35-44). ACM. New York, NY, USA. doi: 10.1145/1753326.1753333

Bailey, P., Thomas, P., & Hawking, D. (2007). Does brandname influence perceived search result quality? Yahoo!, Google, and WebKumara. *Proceedings of ADCS*, 88-91.

Bailey, P., Craswell, N., White, R. W., Chen, L., Satyanarayana, A., & Tahaghoghi, S. M. (2010). Evaluating search systems using result page context. In *Proceedings of the third symposium on Information interaction in context* (pp. 105-114). ACM. New York, NY, USA. doi: 10.1145/1840784.1840801

Balatsoukas, P., & Ruthven, I. (2012). An eye-tracking approach to the analysis of relevance judgments on the Web: The case of Google search engine. *Journal of the American Society for Information Science and Technology*, *63*(9), 1728-1746. doi: 10.1002/asi.22707

Barsky, E., & Bar-Ilan, J. (2012). The impact of task phrasing on the choice of search keywords and on the search process and success. *Journal of the American Society for Information Science and Technology*, *63*(10), 1987-2005. doi: 10.1002/asi.22654

Basile, V., Peroni, S., Tamburini, F., & Vitali, F. (2015). Topical tags vs non-topical tags: Towards a bipartite classification?. *Journal of Information Science*, 41(4), 486-505. doi: 10.1177/0165551515585283

Belkin, N. J., Kelly, D., Kim, G., Kim, J. Y., Lee, H. J., Muresan, G., ... & Cool, C. (2003). Query length in interactive information retrieval. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval* (pp. 205-212). ACM. New York, NY, USA. doi: 10.1145/860435.860474

Bell, D. J., & Ruthven, I. (2004). Searcher's assessments of task complexity for web searching. In *Advances in Information Retrieval* (pp. 57-71). Springer Berlin Heidelberg. doi: 10.1007/978-3-540-24752-4_5

Bergman, O., Gradovitch, N., Bar-Ilan, J., & Beyth-Marom, R. (2013a). Folder versus tag preference in personal information management. *Journal of the American Society for Information Science and Technology, 64*(10), 1995-2012. doi: 10.1002/asi.22906

Bergman, O., Gradovitch, N., Bar-Ilan, J., & Beyth-Marom, R. (2013b). Tagging personal information: A contrast between attitudes and behavior. *Proceedings of the American Society for Information Science and Technology*, 50(1), 1-8.

Bezdan, E., Kester, L., & Kirschner, P. A. (2013). The influence of node sequence and extraneous load induced by graphical overviews on hypertext learning. *Computers in Human Behavior, 29*, 870-880. doi: 10.1016/j.chb.2012.12.016

Bhattacharya, N., & Gwizdka, J. (2018). Relating Eye-Tracking Measures with Changes in Knowledge on Search Tasks. *arXiv preprint arXiv*:1805.02399.

Bilal, D. (2000). Children's use of the Yahooligans! Web search engine: I. Cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society of Information Science, 51,* 646-665.

Boardman, R., & Sasse, M. A. (2004). Stuff goes into the computer and doesn't come out: a cross-tool study of personal information management. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 583-590). ACM.

Boekhorst, A. K., Kwast, I., & Wevers, D. (2004). *Informatievaardigheden* (3rd ed). Lemma: Utrecht.

Bookmark. (n.d.). In *Merriam-Webster's online dictionary* (11th ed.). Retrieved October 9, 2016, from http://www.merriam-webster.com/dictionary/bookmark

Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation*, *56*(1), 71-90. doi: 10.1108/EUM0000000007110

Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for information Science and Technology*, *54*(10), 913-925. doi: 10.1002/asi.10286

Borlund, P., & Dreier, S. (2014). An investigation of the search behaviour associated with Ingwersen's three types of information needs. *Information Processing & Management*, *50*(4), 493-507. doi: 10.1016/j.ipm.2014.03.001

Borlund, P., & Schneider, J. W. (2010). Reconsideration of the simulated work task situation: A context instrument for evaluation of information retrieval interaction. In *Proceedings of the third symposium on Information interaction in context* (pp. 155-164). ACM. doi: 10.1145/1840784.1840808

Brand-Gruwel, S., & Wopereis, I. (2006). Integration of the information problem-solving skill in a curriculum: The effects of learning with authentic tasks. *Technology, Instruction, Cognition and Learning, 4*, 243–263.

Brand-Gruwel, S., Kammerer, Y., Meeuwen, L., & Van Gog, T. (2017). Source evaluation of domain experts and novices during Web search. *Journal of Computer Assisted Learning, 33*(3), 234-251. doi: 10.1111/jcal.12162

Brand-Gruwel, S., Wopereis, I.G.J.H., & Vermetten, Y. (2005). Information problem solving by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behavior, 21*, 487–508. doi: 10.1016/j.chb.2004.10.005

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using Internet. *Computers & Education, 53*(4), 1207-1217. doi: 10.1016/j.compedu.2009.06.004

Brannon, L. A., Tagler, M. J., & Eagly, A. H. (2007). The moderating role of attitude strength in selective exposure to information. *Journal of Experimental Social Psychology, 43*(4), 611–617. doi: 10.1016/j.jesp.2006.05.001

Brennan, K., Kelly, D., & Arguello, J. (2014). The effect of cognitive abilities on information search for tasks of varying levels of complexity. In *Proceedings of the 5th Information Interaction in Context Symposium* (pp. 165-174). ACM. doi: 10.1145/2637002.2637022

Broder, A. (2002). A taxonomy of web search. *In ACM Sigir forum* (Vol. 36, No. 2, pp. 3-10). ACM, New York. doi: 10.1145/792550.792552

Bruce, H. (2005). Personal, Anticipated Information Need. *Information Research, 10*(3) paper 332.

Bruce, H., Jones, W., & Dumais, S. (2004). Keeping and re-finding information on the web: What do people do and what do they need?. *Proceedings of the American Society for Information Science and Technology*, *41*(1), 129-137. doi: 10.1002/meet.1450410115

Bush, V. (1945). As we may think. *The Atlantic Monthly, 176*(1), 101-108.

Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information processing & management*, *31*(2), 191-213. doi: 10.1016/0306-4573(95)80035-R

Campagnoni, F. R., & Ehrlich, K. (1989). Information retrieval using a hypertext-based help system. *ACM Transactions on Information Systems (TOIS), 7*, 271-291. ACM. New York, NY, USA. doi: 10.1145/65943.65948

Capra III, R. G., & Pérez-Quinones, M. A. (2005). Using web search engines to find and refind information. *IEEE Computer, 38*(10), 36-42. doi: 10.1109/MC.2005.355

Case, D. (2007). *Looking for information. A survey of research on information needs, seeking and behavior*. Bingley, UK: Emerald.

Chen, J. M., Chen, M. C., & Sun, Y. S. (2014). A tag based learning approach to knowledge acquisition for constructing prior knowledge and enhancing student reading comprehension. *Computers & Education, 70*, 256-268. doi: 10.1016/j.compedu.2013.09.002

Chen, H., & Dumais, S. (2000). Bringing order to the web: Automatically categorizing search results. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 145-152). ACM. New York, NY, USA. doi: 10.1145/332040.332418

Chen, S. Y., & Macredie, R. (2010). Web-based interaction: A review of three important human factors. *International Journal of Information Management, 30*(5), 379-387. doi: 10.1016/j.ijinfomgt.2010.02.009

Choi, Y. (2010). Effects of contextual factors on image searching on the Web. *Journal of the American Society for information Science and Technology*, *61*(10), 2011-2028. doi: 10.1002/asi.21386

Civan, A., Jones, W., Klasnja, P., & Bruce, H. (2008). Better to organize personal information by folders or by tags?: The devil is in the details. *Proceedings of the American Society for Information Science and Technology*, *45*(1), 1–13. doi: 10.1002/meet.2008.1450450214

Clark, R. E., Kirschner, P. A., & Sweller, J. (2012). Putting students on the path to learning: The case for fully guided instruction. *American Educator, 36*(1), 6-11.

Conklin, J. (1987). Hypertext: an introduction and survey. IEEE Computer, 20, 17-41. doi: 10.1109/MC.1987.1663693

Cutrell, E., & Guan, Z. (2007). What are you looking for?: an eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 407-416). ACM, USA. doi: 10.1145/1240624.1240690.

Dee-Lucas, D., & Larkin, J. H. (1999). Hypertext segmentation and goal compatibility: Effects on study strategies and learning. *Journal of Educational Multimedia and Hypermedia, 8*, 279-313.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., ... & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences, 113*(3), 554-559. doi: 10.1073/pnas.1517441113

DeStefano, D., & LeFevre, J. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior, 23*, 1616-1641. doi: 10.1016/j.chb.2005.08.012

De Vries, B., Van der Meij, H., & Lazonder, A. (2008). Supporting reflective web searching in elementary schools. *Computers in Human Behavior, 24*, 649-665. doi: 10.1016/j.chb.2007.01.021

Dourish, P., Edwards, W. K., LaMarca, A., Lamping, J., Petersen, K., Salisbury, M., ... & Thornton, J. (2000). Extending document management systems with user-specific active properties. *ACM Transactions on Information Systems (TOIS), 18*(2), 140-170.

Duggan, G. B., & Payne, S. J. (2008). Knowledge in the head and on the web: Using topic expertise to aid search. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 39-48). ACM. New York, NY, USA. Florence, Italy. doi: 10.1145/1357054.1357062

Dumais, S. T., & Landauer, T. K. (1983). Using examples to describe categories. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 112-115). Boston, MA: ACM Press.

Eggemeier, F. T., & Wilson, G. F. (1991). Performance-based and subjective assessment of workload. In D. L. Damos (Ed.), *Multiple task performance* (pp. 217–278). London: Taylor & Francis.

Eisenberg, M. B., & Berkowitz, R. E. (1990). *Information problem-solving: The big six skills approach to library and information skills instruction*. Norwood, NJ: Ablex.

Fastrez, P., & Jacques, J. (2015). Managing References by Filing and Tagging. In *International Conference on Human Interface and the Management of Information* (pp. 291-300). Springer, Cham. doi: 10.1007/978-3-319-20612-7_28

Finley, K. (2015, April 15). EU Formally Accuses Google of Antitrust Violations. *Wired*. Retrieved from: http://www.wired.com/2015/04/eu-google/

Fisher, K. E., Erdelez, S., & McKechnie, L. (Eds.). (2005). *Theories of information behavior*. Information Today, Inc.

Fischer, P., & Greitemeyer, T. (2010). A new look at selective-exposure effects: An integrative model. *Current Directions in Psychological Science, 19*(6), 384-389. doi: 0.1177/0963721410391246

Ford, N., Miller, D., & Moss, N. (2001). The role of individual differences in Internet searching: An empirical study. *Journal of the American Society for Information Science and Technology, 52*, 1049–1066. doi: 10.1002/asi.1165

Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction, 21*, 220-231. doi: 10.1016/j.learninstruc.2010.02.005

Golder, S. A., & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of information science*, *32*(2), 198-208. doi: 10.1177/0165551506062337

Granka, L. A., Joachims, T., & Gay, G. (2004). Eye-tracking analysis of user behavior in WWW search. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 478-479). ACM. doi: 10.1145/1008992.1009079

Guan, Z., & Cutrell, E. (2007). An eye tracking study of the effect of target rank on web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 417-420). ACM. doi: 10.1145/1240624.1240691

Guy, M. & Tonkin, E. (2006). Folksonomies: Tidying up tags? *D-Lib Magazine, 12*(1). Retrieved July 11, 2009 from http://www.dlib.org/dlib/january06/guy/01guy.html

Gwizdka, J., & Spence, I. (2006). What can searching behavior tell us about the difficulty of information tasks? A study of Web navigation. *Proceedings of the American Society for Information Science and Technology*, *43*(1), 1-22. doi: 10.1002/meet.14504301167

Hargittai, E., Fullerton, L., Menchen-Trevino, E., & Thomas, K. Y. (2010). Trust online: Young adults' evaluation of web content. *International Journal of Communication, 4*, 27, 468-494.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, No. 9, pp. 904-908). Sage Publications. doi: 10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, *52*, 139-183. doi: 10.1016/S0166-4115(08)62386-9

Head, A. J., & Eisenberg, M. B. (2011). How college students use the Web to conduct everyday life research. *First Monday, 16* (4), Retrieved at April 30, 2013, from http://firstmonday.org/article/view/3484/2857. doi: 10.5210/fm.v16i4.3484

Heckner, M., Heilemann, M., & Wolff, C. (2009). Personal Information Management vs. Resource Sharing: Towards a Model of Information Behavior in Social Tagging Systems. In *Proceedings of the Third International AAAI Conference on Weblogs and Social Media*.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of memory and language*, *32*(6), 717-742. doi: 10.1006/jmla.1993.1036

Hilbert, T. S., & Renkl, A. (2009). Learning how to use a computer-based concept-mapping tool: Self-explaining examples helps. *Computers in Human Behavior*, *25*(2), 267-274. doi: 10.1016/j.chb.2008.12.006

Hochmair, H. H., & Luttich, K. (2006). An Analysis of the Navigation Metaphor – and Why It Works for the World Wide Web. *Spatial Cognition and Computation, 6*, 235-278. doi: 10.1207/s15427633scc0603_3

Höchstötter, N., & Lewandowski, D. (2009). What users see – Structures in search engine results pages. *Information Sciences, 179* (12), 1796-1812. doi: 10.1016/j.ins.2009.01.028

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Hupfer, M., & Detlor, B. (2006). Gender and Web information seeking: A self-concept orientation model. *Journal of the American Society for Information Science and Technology, 57*, 1105–1115. doi: 10.1002/asi.20379

Huvila, I. (2013). In Web search we trust?: Articulation of the cognitive authorities of Web searching. *Information Research*, 18(1).

Ingwersen, P. (2000). Users in context. In *Lectures on information retrieval* (pp. 157-178). Springer Berlin Heidelberg. doi: 10.1007/3-540-45368-7_8

Jansen, B. J., Booth, D. L., & Spink, A. (2008). Determining the informational, navigational, and transactional intent of Web queries. *Information Processing & Management, 44*(3), 1251-1266. doi: 10.1016/j.ipm.2007.07.015

Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing and Management, 42*, 248-263. doi: 10.1016/j.ipm.2004.10.007

Jansen, B. J., Spink, A., & Koshman, S. (2007). Web searcher interaction with the Dogpile.com metasearch engine. *Journal of the American Society for Information Science and Technology*, *58*(5), 744-755. doi: 10.1002/asi.20555

Järvelin, K., Vakkari, P., Arvola, P., Baskaya, F., Järvelin, A., Kekäläinen, J., ... & Sormunen, E. (2015). Task-Based Information Interaction Evaluation: The Viewpoint of Program Theory. *ACM Transactions on Information Systems (TOIS)*, *33*(1), 3. doi: 10.1145/2699660

Jones, W. (2007). Personal information management. *Annual review of information science and technology*, *41*(1), 453-504. doi: 10.1002/aris.2007.1440410117

Juvina, I., & Van Oostendorp, H. V. (2006). Individual differences and behavioral metrics involved in modeling web navigation. *Universal Access in the Information Society, 4*, 258–269. doi: 10.1007/s10209-005-0007-7

Kalyuga, S. (2007). Expertise reversal effect and its implication for learner-tailored instruction. *Educational Psychology Review, 19*, 509-539. doi: 10.1007/s10648-007-9054-3

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38(1)*, 23-31. doi: 10.1207/S15326985EP3801_4

Kammerer, Y., & Beinhauer, W. (2010). Gaze-based web search: the impact of interface design on search result selection. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (pp. 191-194). ACM. doi: 10.1145/1743666.1743713

Kammerer, Y., & Bohnacker, M. (2012). Children's web search with Google: the effectiveness of natural language queries. In *proceedings of the 11th International Conference on Interaction Design and Children* (pp. 184-187). ACM. New York, NY, USA. doi: 10.1145/2307096.2307121

Kammerer, Y., & Gerjets, P. (2012). Effects of search interface and Internet-specific epistemic beliefs on source evaluations during web search for medical information: An eye-tracking study. *Behaviour & Information Technology, 31* (1), 83-97. doi: 10.1080/0144929X.2011.599040

Kammerer, Y., & Gerjets, P. (2014). The role of search result position and source trustworthiness in the selection of web search results when using a list or a grid interface. *International Journal of Human-Computer Interaction*, *30*(3), 177-191. doi: 10.1080/10447318.2013.846790

Kellar, M., Watters, C., & Sheperd, M. (2007). A field study characterizing web-based information-seeking tasks. *Journal of American Society for Information Science and Technology, 58* (7), 999–1018. doi: 10.1002/asi.20590

Kelly, D., Arguello, J., Edwards, A., & Wu, W. C. (2015). Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework. In *Proceedings of the 2015 International Conference* *on The Theory of Information Retrieval* (pp. 101-110). ACM. New York, NY, USA. doi: 10.1145/2808194.2809465

Kidd, A. (1994). The marks are on the knowledge worker. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 186-191). ACM. New York, NY, USA. doi: 10.1145/191666.191740

Kim, H., & Hirtle, S. C. (1995). Spatial metaphors and disorientation in hypertext browsing. *Behaviour & Information Technology, 14*(4), 239-250. doi: 10.1080/01449299508914637

Kim, J., Kazai, G., & Zitouni, I. (2013). Relevance dimensions in preference-based IR evaluation. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval* (pp. 913-916). ACM. New York, NY, USA.

Kim, K. S., & Allen, B. (2002). Cognitive and task influences on web searching behavior. *Journal of the American Society for Information Science and Technology, 53*(2), 109–119. doi: 10.1002/asi.10014

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist, 41*(2), 75-86. doi: 10.1207/s15326985ep4102_1

Kuhlthau, C. (2004). *Seeking meaning: A process approach to library and information services* (2nd ed). Westport: Libraries Unlimited.

Lansdale, M. W. (1988). The psychology of personal information management. *Applied Ergonomics, 19*(1), 55-66.

Large, A., & Beheshti, J. (2000). The web as a classroom resource: Reaction from the users. *Journal of the American Society of Information Science, 51*, 1069-1080. doi: 10.1002/1097-4571(2000)9999:9999%3C::AID-ASI1017%3E3.0.CO;2-W

Large, A., Beheshti, J., & Rahman, T. (2002). Gender differences in collaborative web searching behavior: an elementary school study. *Information Processing & Management, 38*, 427–443. doi: 10.1016/S0306-4573(01)00034-6

Lee, C. S., Goh, D. H. L., Razikin, K., & Chua, A. Y. (2009). Tagging, sharing and the influence of personal experience. *Journal of Digital Information*, 10(1). Retrieved from https://journals.tdl.org/jodi/index.php/jodi/article/view/275

Lee, H., & Pang, N. (2017). Understanding the effects of task and topical knowledge in the evaluation of websites as information patch. *Journal of Documentation*. doi: 10.1108/JD-04-2017-0050

Leuthold, S., Schmutz, P., Bargas-Avila, J. A., Tuch, A. N., & Opwis, K. (2011). Vertical versus dynamic menus on the world wide web: Eye tracking study measuring the influence of menu design and task complexity on user performance and subjective preference. *Computers in Human Behavior, 27*, 459-472. doi: 10.1016/j.chb.2010.09.009

Lewandowski, D. (2012). Credibility in Web search engines. In S. Apostel & M. Folk (eds.) *Online credibility and digital ethos: Evaluating computer-mediated communication*, 131-146. Hershey, PA: IGI Global. doi: 10.4018/978-1-4666-2663-8.ch008

Lewandowski, D. (2015). Evaluating the retrieval effectiveness of Web search engines using a representative query sample. *Journal of the Association for Information Science and Technology*. doi: 10.1002/asi.23304

Lewandowsky, S., Oberauer, K., & Gignac, G. E. (2013). NASA faked the moon landing—therefore, (climate) science is a hoax: An anatomy of the motivated rejection of science. *Psychological science, 24*(5), 622-633. doi: 10.1177/0956797612457686

Linstone, H. A., & Turoff, M. (1975). *The Delphi method: Techniques and applications*. Reading, MA: Addison-Wesley.

Linstone, H. A., & Turoff, M. (2011). Delphi: A brief look backward and forward. *Technological Forecasting and Social Change*, *78*(9), 1712-1719. doi: 10.1016/j.techfore.2010.09.011

Liu, J., Gwizdka, J., Liu, C., & Belkin, N. J. (2010). Predicting task difficulty for different task types. *Proceedings of the American Society for Information Science and Technology*, *47*(1), 1-10. doi: 10.1002/meet.14504701173

Loftus, G. R. (1981). Tachistoscopic simulations of eye fixations on pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(5), 369. doi: 10.1037/0278-7393.7.5.369

Lohmann, S., Ziegler, J., & Tetzlaff, L. (2009). Comparison of tag cloud layouts: Task-related performance and visual exploration. In *Human-Computer Interaction–INTERACT 2009* (pp. 392-404). Springer Berlin Heidelberg. doi: 10.1007/978-3-642-03655-2_43

Lorenzen, M. (2002). The land of confusion? High school students and their use of the World Wide Web for research. *Research strategies, 18*(2), 151–163.

Machill, M., Neuberger, C., Schweiger, W., & Wirth, W. (2004). Navigating the Internet: A study of German-language search engines. *European Journal of Communication, 19*, 321-347. doi: 10.1177/0267323104045258

Madrid, R. I., Van Oostendorp, H., & Puerta Melguizo, M. C. (2009). The effects of the number of links and navigation support on cognitive load and learning with hypertext: The mediating role of reading order. Computers in Human Behavior, 25, 66-75. doi: 10.1016/j.chb.2008.06.005

Maglio, P. P., & Matlock, T. (1998). Metaphors we surf the web by. In *Workshop on Personalized and Social Navigation in Information Space*.

MaKinster, J. G., Beghetto, R. A., & Plucker, J. A. (2002). Why can't I find Newton's third law? Case studies of students' use of the web as a science resource. *Journal of Science Education and Technology, 11*, 155-172. doi: 10.1023/A:1014617530297

Malone, T. W. (1983). How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Office Information Systems, 1*, 99-112. doi: 10.1145/357423.357430

Mankoff, J., Rode, J. A., & Faste, H. (2013). Looking past yesterday's tomorrow: using futures studies methods to extend the research horizon. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1629-1638). ACM. New York, USA. doi: 10.1145/2470654.2466216

Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. *Journal of the American Society for Information Science*, *40*(1), 54-66. doi: 10.1002/(SICI)1097-4571(198901)40:1%3C54::AID-ASI6%3E3.0.CO;2-R

Marshall, C. C. & Bly, S. (2005). Saving and using encountered information: Implications for electronic periodicals. *Proceedings of CHI '05*, 111-120. doi: 10.1145/1054972.1054989

Martens, R. L. (2007). *Positive learning met multimedia. Onderzoeken, toepassen & generaliseren*. Inaugural speech. Heerlen, the Netherlands: Open Universiteit Nederland.

McCrory Wallace, R., Kupperman, J., Krajcik, J., & Soloway, E. (2000). Science on the Web: Students online in a sixth-grade classroom. *The Journal of the Learning Sciences*, *9*(1), 75-104. doi: 10.1207/s15327809jls0901_5

Melenhorst, M., & van Setten, M. (2007). Usefulness of tags in providing access to large information systems. In *2007 IEEE International Professional Communication Conference* (pp. 1-9). IEEE. Seattle, WA, USA. doi: 10.1109/IPCC.2007.4464070

Minetou, C., Chen, S., & Liu, X. (2008). Investigation of the use of navigation tools in web-based learning: A data mining approach. *International Journal of Human-Computer Interaction, 24*, 48-67. doi: 10.1080/10447310701771480

Mobrand, K. A., & Spyridakis, J. H. (2007). Explicitness of local navigational links: comprehension, perceptions of use, and browsing behavior. *Journal of Information Science, 33*, 41-61. doi: 10.1177/0165551506068144

Monchaux, S., Amadieu, F., Chevalier, A., & Mariné, C. (2015). Query strategies during information searching: Effects of prior domain knowledge and complexity of the information problems to be solved. *Information Processing and Management, 51*(5), 557–569. doi: 10.1016/j.ipm.2015.05.004

Moore, P. (1995). Information problem solving: A wider view of library skills. *Contemporary educational psychology, 20*(1), 1-31. doi: 10.1006/ceps.1995.1001

Mosenthal, P. B. (1998). Defining prose task characteristics for use in computer-adaptive testing and instruction. *American Educational Research Journal, 35*, 269-307. doi: 10.3102/00028312035002269

Nyström, M., Andersson, R., Holmqvist, K., & Van De Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior research methods, 45*(1), 272-288. doi: 10.3758/s13428-012-0247-4

Oh, K. E., & Belkin, N. J. (2014). Understanding what personal information items make categorization difficult. *Proceedings of the Association for Information Science and Technology, 51*(1), 1-3. doi: 10.1002/meet.2014.14505101139

Pak, R., Pautz, S., & Iden, R. (2007). Information organization and retrieval: A comparison of taxonomical and tagging systems. *Cognitive Technology, 12*(1), 31-44.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*, *12*(3), 801-823. doi: 10.1111/j.1083-6101.2007.00351.x

Pitkow, J. E., & Kehoe, C. M. (1996). Emerging trends in the WWW user population. *Communications of the ACM*, *39*(6), 106-108. ACM. New York, NY, USA. doi: 10.1145/228503.228525

Puerta Melguizo, M. C., Vidya, U., & Van Oostendorp, H. (2012). Seeking information online: the influence of menu type, navigation path complexity and spatial ability on information gathering tasks. *Behaviour & Information Technology, 31*, 59-70. doi: 10.1080/0144929X.2011.602425

Purcell, K., Rainie, L., Heaps, A., Buchanan, J., Friedrich, L., Jacklin, A., Chen, C., & Zickuhr, K. (2012). *How teens do research in the digital world.* Washington, United States of America: Pew Research Center's Internet & American Life Project.

Qiu, L. (1993). Analytical Searching vs. Browsing in Hypertext Information Retrieval Systems. *Canadian Journal of Information and Library Science*, *18*(4), 1-13.

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from: http://www.R-project.org/

Rader, E., & Wash, R. (2008). Influences on tag choices in del.icio.us. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (pp. 239-248). ACM. New York, NY, USA.

Rogers, D., & Swan, K. (2004). Self-regulated learning and Internet searching. *Teachers College Record, 106*(9), 1804-1824. doi: 10.1111/j.1467-9620.2004.00406.x

Rose, D. E., & Levinson, D. (2004, May). Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web* (pp. 13-19). ACM. New York, NY, USA. doi: 10.1145/988672.988675

Rouet, J. F. (2003). What was I looking for? The influence of task specificity and prior knowledge on students search strategies in hypertext. *Interacting with Computers, 15*(3), 409–428. doi: 10.1016/S0953-5438(02)00064-4

Rouet, J., Ros, C., Goumi, A., Macedo-Rouet, M., & Dinet, J. (2011). The influence of surface and deep cues on primary and secondary school students' assessment of relevance in Web menus. *Learning and Instruction*, *21*, 205-219. doi: 10.1016/j.learninstruc.2010.02.007

Roy, M., & Chi, M. T. H. (2003). Gender Differences in Patterns of Searching the Web. *Journal of Educational Computing Research, 29*, 335–348. doi: 10.2190/7BR8-VXA0-07A7-8AVN

Saastamoinen, M., Kumpulainen, S., & Järvelin, K. (2012). Task complexity and information searching in administrative tasks revisited. In *Proceedings of the 4th Information Interaction in Context Symposium* (pp. 204-213). ACM. doi: 10.1145/2362724.2362759

Salmerón, L., & Kammerer, Y. (2012). Selecting pages from Google to learn about a controversial topic: the role of epistemic belief. In E. de Vries & K. Scheiter (Eds.), *Proceedings EARLI Special Interest Group Text and Graphics: Staging knowledge and experience: How to take advantage of representational technologies in education and training?*, 181-183.

Salmerón, L., Kammerer, Y., & García-Carrión, P. (2013). Searching the Web for conflicting topics: Page and user factors. *Computers in Human Behavior, 29*(6), 2161-2171. doi: 10.1016/j.chb.2013.04.034

Salomon, G., & Almog, T. (1998). Educational Psychology and Technology: A Matter of Reciprocal Relations. *Teachers College Record, 100*(2), 222-241.

Sanchiz, M., Chevalier, A., & Amadieu, F. (2017). How do older and young adults start searching for information? Impact of age, domain knowledge and problem complexity on the different steps of information searching. *Computers in Human Behavior, 72*, 67-78. doi: 10.1016/j.chb.2017.02.038

Savenkov, D., Braslavski, P., & Lebedev, M. (2011). Search snippet evaluation at yandex: lessons learned and future directions. In *Multilingual and Multimodal Information Access Evaluation* (pp. 14-25). Springer Berlin Heidelberg. doi: 10.1007/978-3-642-23708-9_4

Scheiter, K., & Gerjets, P. (2007). Learner control in hypermedia environments. *Educational Psychology Review, 19*, 285-307. doi: 10.1007/s10648-007-9046-3

Selberg, E., & Etzioni, O. (2000). On the Instability of Web Search Engines. *RIAO*, 223-236. LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE. Paris, France.

Shen, S. T., & Prior, S. D. (2013). My favorites (bookmarks) schema: one solution to online information storage and retrieval. In *Proceedings of the 2013 International Conference on Information Systems and Design of Communication* (pp. 33-40). ACM. New York, NY, USA. doi: 10.1145/2503859.2503865

Shirky, C. (2005). Ontology is overrated: Categories, links and tags. Retrieved July 12, 2009, from http://shirky.com/writings/ontology_overrated.html.

Silverstein, C., Henzinger, M., Marais, H., & Moricz, M. (1998). *Analysis of a very large AltaVista query log.* Palo Alto, California, United States of America: Systems Research Center.

Singer, G., Norbisrath, U., & Lewandowski, D. (2012a). Ordinary search engine users carrying out complex search tasks. *Journal of Information Science,* 1-13. doi: 10.1177/0165551512466974

Singer, G., Pruulmann-Vengerfeldt, P., Norbisrath, U., & Lewandowski, D. (2012b). The relationship between Internet user type and user performance when carrying out simple vs. complex search tasks. *First Monday, 17* (6), Retrieved at August 26, 2015, from http://firstmonday.org/article/view/3960/3245. doi: 10.5210/fm.v17i6.3960

Social bookmarking. (2015, November 14). In *Wikipedia, The Free Encyclopedia*. Retrieved December 13, 2015, from https://en.wikipedia.org/w/index.php?title=Social_bookmarking&oldid=690533455

Song, Y., Ma, H., Wang, H., & Wang, K. (2013). Exploring and exploiting user search behavior on mobile and tablet devices to improve search relevance. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 1201-1212). ACM. Rio de Janeiro, Brazil. doi: 10.1145/2488388.2488493

Špiranec, S., & Ivanjko, T. (2013). Experts vs. novices tagging behavior: an exploratory analysis. *Procedia-Social and Behavioral Sciences, 73*, 456-459. doi: 10.1016/j.sbspro.2013.02.076

Stadtler, M., & Bromme, R. (2008). Effects of the metacognitive computer-tool met.a.ware on the websearch of laypersons. *Computers in Human Behavior, 24(3)*, 716-737. doi: 10.1016/j.chb.2007.01.023

Stanney, K. M., & Salvendy, G. (1995). Information visualization; assisting low spatial individuals with information access tasks through the use of visual mediators. *Ergonomics, 38*, 1184-1198. doi: 10.1080/00140139508925181

Sterling, G. (2015). It's Official: Google Says More Searches Now On Mobile Than On Desktop. *Search Engine Land*. Retrieved on June 2nd, 2018 from: https://searchengineland.com/its-official-google-says-more-searches-now-on-mobile-than-on-desktop-220369

Sunstein, C. R., & Vermeule, A. (2009). Conspiracy theories: Causes and cures. *Journal of Political Philosophy, 17*, 202–227. doi: 10.1111/j.1467-9760.2008.00325.x

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science, 12*(2), 257-285. doi: 10.1207/s15516709cog1202_4

Symantec. (2010). Survey reveals poor availability of digital evidence brings Legal Process to a halt across EMEA. Retrieved December 19, 2015 from: http://www.symantec.com/en/uk/about/news/release/article.jsp?prid=20100907_01

Tauscher, L., & Greenberg, S. (1997). How people revisit webpages: Empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies, 47(1)*, 97-137. doi: 10.1006/ijhc.1997.0125

Teevan, J. (2008). How people recall, recognize, and reuse search results. *ACM Transactions on Information Systems (TOIS)*, *26*(4), 19. ACM. New York, NY, USA. doi: 10.1145/1402256.1402258

Teevan, J., Adar, E., Jones, R., & Potts, M. A. (2007). Information re-retrieval: repeat queries in Yahoo's logs. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 151-158). ACM. New York, NY, USA. doi: 10.1145/1277741.1277770

Tobii 1750 [Apparatus]. (2003). Danderyd, Sweden: Tobii Technology AB.

Tobii Studio [Software]. (2007). Danderyd, Sweden: Tobii Technology AB.

Tombros, A., & Sanderson, M. (1998). Advantages of query biased summaries in information retrieval. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 2-10). ACM. doi: 10.1145/290941.290947

Toms, E. G., O'Brien, H., Mackenzie, T., Jordan, C., Freund, L., Toze, S., ... & Macnutt, A. (2008). Task effects on interactive search: The query factor. In *Focused access to XML documents* (pp. 359-372). Springer Berlin Heidelberg. doi: 10.1007/978-3-540-85902-4_31

Trattner, C., Lin, Y. L., Parra, D., Yue, Z., Real, W., & Brusilovsky, P. (2012). Evaluating tag-based information access in image collections. In *Proceedings of the 23rd ACM conference on Hypertext and social media* (pp. 113-122). ACM. New York, NY, USA. doi: 10.1145/2309996.2310016

Vakkari, P. (1999). Task complexity, problem structure and information actions: Integrating studies on information seeking and retrieval. *Information processing & management*, *35*(6), 819-837. doi: 10.1016/S0306-4573(99)00028-X

Van Deursen, A. J. A. M., & Van Dijk, J. A. G. M. (2008) *Digitale vaardigheden van Nederlandse burgers. Een prestatiemeting van operationele, formele, informatie en strategische vaardigheden bij het gebruik van overheidswebsites [Digital skills of Dutch citizens. A performance measurement of operational, formal, information and strategic skills in the use of government websites]*. (Doctoral dissertation). Enschede, the Netherlands: University of Twente.

Van Merriënboer, J. J. G., & Kirschner, P. A. (2012). *Ten steps to complex learning (2nd ed.)*. New York: Taylor & Francis.

Van Strien, J. L. H., Brand-Gruwel, S., & Boshuizen, H. P. A. (2014). Dealing with conflicting information from multiple nonlinear texts: Effects of prior attitudes. *Computers in Human Behavior, 32*, 101-111. doi: 10.1016/j.chb.2013.11.021

Veldhorst, G., Frerejean, J., Van Strien, J. & Brand-Gruwel, S. (submitted). The impact of prior attitudes and prior knowledge on the problem formulation stage of information problem solving.

Voit, K., Andrews, K., & Slany, W. (2009). Why personal information management (PIM) technologies are not widespread. In *PIM Workshop, ASIS&T* (Vol. 2009, pp. 60-64).

Voit, K., Andrews, K., & Slany, W. (2012). Tagging might not be slower than filing in folders. *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA '12*, 2063. doi: 10.1145/2212776.2223753

Walhout, J., Brand-Gruwel, S., Jarodzka, H., Van Dijk, M., De Groot, R. & Kirschner, P. A. (2015). Learning and Navigating in Hypertext: Navigational Support by Hierarchical Menu or Tag Cloud? *Computers in Human Behavior, 46,* 218-227. doi: 10.1016/j.chb.2015.01.025

Walhout, J., Jarodzka, H., Van Strien, J.L.H., & Brand-Gruwel, S. (submitted). To tag or not to tag? How to support organizing and classifying bookmarks from the web?

Walhout, J., Oomen, P., Jarodzka, H., & Brand-Gruwel, S. (2017). Effects of task complexity on online search behavior of adolescents. *Journal of the Association for Information Science and Technology, 68*(6), 1449-1461. doi: 10.1002/asi.23782

Walker, G., & Janes, J. (1999). *Online retrieval: A dialogue of theory and practice (2nd ed.)*. Englewood, CO: Libraries Unlimited.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2008). Information problem solving: A review of problems students encounter and instructional solutions. *Computers in Human Behavior, 24* (3), 623-648. doi: 10.1016/j.chb.2007.01.030

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Computers & Education*, *52*(1), 234-246. doi: 10.1016/j.compedu.2008.08.003

Wash, R., & Rader, E. (2007). Public bookmarks and private benefits: An analysis of incentives in social computing. *Proceedings of the American Society for Information Science and Technology*, *44*(1), 1-13. doi: 10.1002/meet.1450440240

Westerwick, A. (2013). Effects of sponsorship, Web site design, and Google ranking on the credibility of online information. *Journal of Computer-Mediated Communication*, 18(2), 80-97. doi:10.1111/jcc4.12006

White, R. W. (2014). Belief dynamics in Web search. *Journal of the Association for Information Science and Technology*, *65*(11), 2165-2178. doi: 10.1002/asi.23128

White, R. W., Dumais, S. T., & Teevan, J. (2009). Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*. 132-141. ACM. doi: 10.1145/1498759.1498819

White, R. W., Jose, J. M., & Ruthven, I. (2003). A task-oriented study on the influencing effects of query-biased summarisation in web searching. *Information Processing & Management*, *39*(5), 707-733. doi: 10.1016/S0306-4573(02)00033-X

Whittaker, S. (2011). Personal information management: from information consumption to curation. *Annual Review of Information Science and Technology, 45*(1), 1-62. doi: 10.1002/aris.2011.1440450108

Whittaker, S., & Sidner, C. (1996). Email overload: exploring personal information management of email. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 276-283). ACM. doi: 10.1145/238386.238530

Wichowski, A. (2009). Survival of the fittest tag: Folksonomies, findability, and the evolution of information organization. *First Monday, 14*(5). Retrieved July 11, 2009, from http://firstmonday.org/ojs/index.php/fm/article/view/2447/2175

Wiebe, E. N., Roberts, E., & Behrend, T. S. (2010). An examination of two mental workload measurement approaches to understanding multimedia learning. *Computers in Human Behavior*, *26*(3), 474-481. doi: 10.1016/j.chb.2009.12.006

Wildemuth, B. M. (2004). The effects of domain knowledge on search tactic formulation. *Journal of the Association for Information Science and Technology, 55*(3), 246-258. doi: 10.1002/asi.10367

Wildemuth, B. M., & Freund, L. (2012). Assigning search tasks designed to elicit exploratory search behaviors. In *Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval* (p. 4). ACM. doi: 10.1145/2391224.2391228

Wildemuth, B. M., Freund, L., & G. Toms, E. (2014). Untangling search task complexity and difficulty in the context of interactive information retrieval studies. *Journal of Documentation*, *70*(6), 1118-1140.

Wolf, S. E., Brush, T., & Saye, J. (2003). Using Information Problem-Solving Model as a Metacognitive Scaffold for Multimedia-Supported Information-Based Problems. *Journal of Research on Technology in Education, 35*(3), 312-342. doi: 10.1080/15391523.2003.10782389

Wright, A. (2007). *Glut. Mastering Information Through the Ages*. Washington D.C., USA: Joseph Henry Press.

Zumbach, J., & Mohraz, M. (2008). Cognitive load in hypermedia reading comprehension: Influence of text type and linearity. *Computers in Human Behavior*, *24*(3), 875-887. doi: 10.1016/j.chb.2007.02.015

Summary

When you would have given teenagers 25 years ago the homework assignment to find out how changes in our way of living could alter the process of global warming or in which way emancipation has developed is the past century, they would have tried to think about it, ask someone or started to look into large (book) encyclopedias. Nowadays, the answer to same assignments is just a few clicks away. However, the process of finding the requested information on the Internet comes with several challenges. For instance, students often do not know how to formulate an appropriate search query, how to judge the validity of websites, or how to question the sources. Instead, the choice for opening a website is mainly guided by the title or summary of the site. These actions are part of Information Problem Solving (IPS). The concept of IPS is defined as a combination of skills needed to access and use information (Brand-Gruwel, Wopereis & Walraven, 2009). The IPS process consists of defining the information problem, searching for information, selecting information, processing information, and presenting information. Mostly students iterate between these parts. In addition to deliberately searching for information, students also often accidentally stumble upon new information on the Internet. Whether it is a result of a deliberate search process or an unintended encounter with new information, students have to decide what to do with the information they deem to be useful and want to re-access at a later time. The question of re-access is also related to the purpose of the search for information. A simple navigational search is less likely to evoke a decision about re-access than a search for information when solving a more complex problem where information has to be compared with previously found information or when information is collected over time. Examples of such complex problems are gathering information on how to deal with a chronic disease or collecting information on the impact of global warming for a TV documentary. Many young people have difficulties in solving such (complex) information problems. However, in contemporary research into IPS skills, the aspect of storing and organizing information has received little attention. The main goal of this dissertation was to fill this research gap.

This dissertation explores ways in which people, and students in particular, can be supported in storing and organizing interesting information from the Internet. The key questions are (1) whether it is better to support people with a hierarchical or a tagging system and (2) whether supporting them with a pre-determined classification scheme is beneficial. Additionally, the presented research in this dissertation addresses the influence of task complexity and the interaction with prior knowledge. Based on the IPS-I model, the studies presented in this dissertation investigate these topics around storing and organizing information in depth.

The first study is presented in Chapter 2 and focuses on the relation between task complexity and search behavior. Evaluation of information during information problem solving processes already starts when trying to select the appropriate search result on a search engine results page (SERP). In this study, we investigated the influence of task complexity on search query formulation, evaluation of search results, and task performance by means of a within-subjects design. Task performance was measured by the accuracy of the answer and time on task. Participants performed three search tasks, namely a fact-finding, cause–

effect, and a controversial topic task. Mosenthal's (1998) framework of prose task characteristics served to assess each task's complexity. This framework is composed of three different dimensions: type of information requested, type of match, and plausibility of distractors. During the actual experiment, participants could search the Internet without any restrictions or time limit. Their perceptual search processes were measured with a combination of log files, eye-tracking data, answer forms, and think-aloud protocols.

While the study in Chapter 2 addresses the search process when searching information freely on the Internet, the study presented in Chapter 3 looks at searching within a given, standardized collection of information sources and how different ways to support navigating through such a collection influence task performance and navigation behavior. For a long time, navigational structures were hierarchical in nature (i.e., following a folder structure). Recently, tag clouds were introduced to navigate large collections of information. A tag cloud is a visual representation of keywords describing the contents of, for example, a website. Tag clouds are considered more flexible and provide a better description of the contents. Although previous research has compared several types of navigational support extensively, few studies exist that compare hierarchical structures with tag clouds directly. In this study, we investigated the effect of a navigation tool, task complexity and the participants gender on navigation behavior, visual processing of the navigation tool and task performance. To study these effects, we constructed two almost identical environments. The only aspect in which these environments differed was the type of navigational support, which was either a hierarchical menu or a tag cloud. Such a collection of information as used in Chapter 3 is compiled during the search process where in the sub process of storing and organizing information this collection builds up. When encountering useful information, students are confronted with the question of how to keep that information so they can use it at a later time.

The study presented in Chapter 4 combined several aspects of the first two studies. As in the first study, participants received a search task for which they collected information on a certain subject (obesity). They were asked to store the information that they judged as useful with a bookmarking tool in which they either used a hierarchical structure or a tag cloud to store the information (like in the second chapter). Besides the effect of the type of organizer, we also investigated whether this storing process is easier when students receive an already existing storage schema vs. when they build such a system by themselves during the search. Thus, we studied how a given vs. no classification support within the two types of organizer (hierarchical vs. tag cloud) influences bookmarking behavior and how these two factors might interact with prior knowledge.

Chapter 5 presents a follow-up study of Chapter 4. In this study, we choose to focus on only the tagging system. This allowed us to better isolate the effect of support with a classification scheme. The aim of this study was to determine the extent in which providing a classification scheme has an effect on experienced task difficulty, tagging behavior and how participants used the given bookmark tool. In this study the participants received a search task on the topic of global warming. They had to store the useful information with a bookmarking tool. The participants were randomly assigned to two conditions: one experimental group received support with an expert defined classification schema; the other group received no classification support and had to construct a classification schema during the task. The expert defined classification schema was constructed with a Delphi procedure.

Each of these chapters addressed one or more of the four main research topics of this thesis. The first main topic of interest is the influence of task complexity, which was considered in Chapters 2 and 3. Both studies found that participants provided less correct answers for the more difficult tasks and thus confirmed the findings of a large body of previous research. Furthermore, task complexity also had an influence on task behavior measurements. An increase in task complexity from a simple fact-finding task to a more complex search task led to an increase in time on task and a decrease in browsing activity. However, as the first study showed, the relationship between task complexity and several measurements is not linear. When a task becomes too complex, no further increase in time on task was observed and the number of search queries decreased. While these two studies confirmed the relationship between task complexity and task performance, they also showed that there are some nuances in this relationship.

Moreover, the influence of type of organizer was studied in Chapters 3 and 4. As stated in both chapters, traditionally, information was organized hierarchically, which has several limitations with regard to flexibility and re-findability. The studies in Chapters 3 and 4 compared the hierarchical and the tagging approaches in two aspects of information problem solving: re-accessing information (Chapter 3) and storing and organizing information (Chapter 4). Both studies showed that neither of the two systems of organizing information could be considered as better. A disadvantage of the tagging system was that participants were hardly familiar with it. Although people express a preference for tags when explicitly asked, they fall back to using a hierarchical system in their actual behavior. Learning something new requires effort, thus they rely on the system they are used to. Nonetheless, participants in both studies who did use a tagging system, did not perform worse compared to participants using a hierarchical system. It can therefore be hypothesized that experience with tagging might change the conclusions drawn thus far.

Chapters 4 and 5 considered the influence of providing support with a classification schema. Having prior knowledge about a topic is advantageous in solving related information problems. Prior knowledge of experts is structured in well-developed cognitive schemata, which are used to compare new information with what is already known. Consequently, experts can process new information much easier than novices as the latter do not have such well-developed internal knowledge structures. One of the most profound effects in both studies was, that participant receiving support with a classification scheme also used more tags that were already present in their prior knowledge structures. Thus, the classification scheme allowed participants to recognize familiar concepts and relate those to the information they wanted to store and organize. Furthermore, the support with a classification scheme led to using classifications more often and a more focused use of classification keywords.

Samenvatting

Wanneer je 25 jaar geleden tieners de huiswerkopdracht had gegeven om erachter te komen hoe veranderingen in onze manier van lezen de opwarming van de aarde kan verminderen of op welke wijze de emancipatie de afgelopen eeuw zich heeft ontwikkeld, dan zouden ze geprobeerd hebben om erover na te denken, het aan iemand te vragen of in encyclopedieën (in boek vorm) te kijken voor een antwoord. Tegenwoordig is het antwoord op dezelfde opdrachten slechts een paar klikken verwijderd. Het vinden van de gevraagde informatie op internet brengt echter verschillende uitdagingen met zich mee. Leerlingen weten bijvoorbeeld vaak niet hoe ze een geschikte zoekopdracht moeten formuleren, hoe ze de validiteit van websites moeten beoordelen, of hoe ze de bronnen kritisch kunnen bekijken. In plaats daarvan wordt de keuze voor het openen van een website voornamelijk bepaald door de titel of samenvatting van de website. Deze acties maken deel uit van *Information Problem Solving* (IPS). Het IPS-concept wordt gedefinieerd als een combinatie van vaardigheden die nodig zijn voor toegang tot en gebruik van informatie (Brand-Gruwel, Wopereis, & Walraven, 2009). Het IPS-proces bestaat uit het definiëren van het informatieprobleem, het zoeken naar informatie, het selecteren van informatie, het verwerken van informatie en het presenteren van informatie. Meestal schakelen studenten tussen deze verschillende vaardigheden. Naast het weloverwogen zoeken naar informatie, stuiten studenten ook vaak per ongeluk op nieuwe informatie op internet. Of het nu het resultaat is van een doelbewust zoekproces of het onbedoeld tegenkomen van nieuwe informatie, studenten moeten beslissen wat ze willen doen met de informatie die ze nuttig vinden en later opnieuw willen kunnen bekijken. De noodzaak voor toegang tot eerder gevonden informatie is ook gerelateerd aan het doel van het zoeken naar informatie. Bij het oplossen van een complexer probleem waarbij informatie moet worden vergeleken met eerder gevonden informatie of wanneer informatie in de loop van de tijd wordt verzameld, is het belangrijk dat de informatie toegankelijk blijft. Voorbeelden van dergelijke complexe problemen zijn het verzamelen van informatie over hoe om te gaan met een chronische ziekte of het verzamelen van informatie over de gevolgen van het broeikaseffect voor een tv-documentaire. Voor veel jongeren is het lastig om dergelijke (complexe) informatieproblemen op te lossen. In hedendaags onderzoek naar IPS-vaardigheden heeft het aspect van bewaren en ordenen van informatie echter weinig aandacht gekregen. Het belangrijkste doel van dit proefschrift is om deze lacune in het wetenschappelijk onderzoek op te vullen.

Dit proefschrift onderzoekt manieren waarop mensen, en studenten in het bijzonder, kunnen worden ondersteund bij het opslaan en organiseren van interessante informatie van internet voor het oplossen van complexe informatieproblemen. De belangrijkste vragen zijn: (1) is het beter om mensen met een hiërarchisch of *tagging*-systeem te ondersteunen en (2) is het ondersteunen met een vooraf bepaald classificatieschema voordelig voor het ordenen van bronnen; (3) in welke mate heeft taakcomplexiteit invloed op de taakprestatie en (4) is er interactie met voorkennis?

De eerste studie is in hoofdstuk 2 beschreven en richt zich op de relatie tussen taakcomplexiteit en zoekgedrag. De evaluatie van informatie tijdens de verschillende stadia van het IPS-proces begint al bij het proberen om het juiste zoekresultaat op een resultatenpagina van een zoekmachine (SERP) te selecteren. In deze studie is de invloed van taakcomplexiteit op de formulering van zoekopdrachten, de evaluatie van zoekresultaten en taakprestaties onderzocht gebruik makend van een 'within-subjects' onderzoeksontwerp. Taakprestaties werden gemeten aan de hand van de nauwkeurigheid van het antwoord en de uitvoeringstijd. De deelnemers voerden drie zoektaken uit: een feiten opdracht, een oorzaak-gevolg opdracht en een zoekopdracht over een controversieel onderwerp. Om de complexiteit van elke taak te beoordelen, is het beoordelingskader van taakkenmerken van Mosenthal (1998) gebruikt. Dit beoordelingskader bestaat uit drie verschillende dimensies: type gevraagde informatie, type overeenkomst en aannemelijkheid van afleiders. Tijdens het daadwerkelijke experiment konden de deelnemers zonder enige beperking of tijdslimiet op internet zoeken. Hun perceptuele zoekprocessen werden gemeten met een combinatie van logbestanden, eye-tracking gegevens, antwoordformulieren en hardop-denkprotocollen.

Terwijl de studie in hoofdstuk 2 het zoekproces behandelt bij het vrij zoeken van informatie op internet, bekijkt het onderzoek in hoofdstuk 3 het zoeken binnen een gegeven verzameling van informatiebronnen. In dit onderzoek wordt van twee verschillende manieren om door zo'n verzameling heen te navigeren, bekeken hoe de taakprestaties en het navigatiegedrag beïnvloed worden. Lange tijd waren navigatiestructuren hiërarchisch van aard (dat wil zeggen: het volgen van een mappenstructuur). Enig tijd geleden zijn 'tagclouds' geïntroduceerd om door grote verzamelingen informatie te kunnen navigeren. Een 'tagcloud' is een visuele weergave van sleutelwoorden die de inhoud van bijvoorbeeld een website beschrijven. 'Tagclouds' worden als flexibeler beschouwd en geven een betere beschrijving van de inhoud. Hoewel eerder onderzoek verschillende soorten navigatie-ondersteuning uitgebreid heeft vergeleken, zijn er maar weinig studies die hiërarchische structuren rechtstreeks vergelijken met 'tagclouds'. In deze studie is het effect van een navigatiemechanisme, taakcomplexiteit en het geslacht van de deelnemers op navigatiegedrag, visuele verwerking van het navigatiemechanisme en taakprestaties onderzocht. Om deze effecten te bestuderen, zijn twee vrijwel identieke omgevingen geconstrueerd. Het enige aspect waarin deze omgevingen verschilden, was het type navigatie ondersteuning: een hiërarchisch menu of een 'tagcloud'. Een dergelijke verzameling van informatie zoals gebruikt in hoofdstuk 3 wordt samengesteld tijdens het zoeken naar informatie. Bij het tegenkomen van nuttige informatie worden studenten geconfronteerd met de vraag hoe ze die informatie moeten bewaren, zodat ze deze op een later tijdstip kunnen gebruiken.

De studie in hoofdstuk 4 combineerde verschillende aspecten van de eerste twee onderzoeken. Net als in de eerste studie ontvingen deelnemers een zoekopdracht waarvoor ze informatie over een bepaald onderwerp (obesitas) verzamelden. Ze werden gevraagd om de informatie die ze nuttig achtten op te slaan met een *bookmarktool* waarin ze een hiërarchische structuur of een 'tagcloud' gebruikten om de informatie op te slaan en te ordenen (zoals in het tweede hoofdstuk). Naast het effect van het structureringstype (hiërarchisch versus 'tagcloud') is ook onderzocht of dit opslagproces gemakkelijker is wanneer studenten een al bestaand classificatieschema ontvangen versus wanneer ze een dergelijk schema zelf bouwen tijdens het zoeken. Zo is onderzocht hoe een gegeven classificatieschema versus geen classificatie ondersteuning binnen de twee structureringstypen invloed heeft op het bookmarkgedrag en hoe deze twee factoren kunnen interacteren met voorkennis.

Hoofdstuk 5 presenteert een vervolgstudie van hoofdstuk 4. In deze studie is gekozen om ons alleen te richten op het tagging-systeem. Hierdoor kon het effect van ondersteuning met een classificatieschema beter worden geïsoleerd. Het doel van deze studie was om te bepalen in hoeverre het aanbieden van een classificatieschema invloed heeft op ervaren moeilijkheid van de taak, tag-gedrag en hoe deelnemers de gegeven bookmarktool hebben gebruikt. In deze studie ontvingen de deelnemers een zoekopdracht over het broeikaseffect. Ze moesten de nuttige informatie opslaan met een bookmarktool. De deelnemers werden willekeurig aan twee condities toegewezen: de experimentele groep kreeg ondersteuning met een door experts gedefinieerd classificatieschema, de andere groep ontving geen classificatieondersteuning en moest tijdens het uitvoeren van de opdracht zelf een classificatieschema opstellen. Het door deskundigen gedefinieerde classificatieschema werd geconstrueerd met behulp van een Delphi-procedure.

Elk van deze hoofdstukken ging in op één of meer van de in totaal vier onderzoekthema's van dit proefschrift. Het eerste onderzoeksthema betreft de invloed van taakcomplexiteit op het oplossen van een informatieprobleem. Dit onderwerp komt in de hoofdstukken 2 en 3 aan de orde. Beide studies vonden dat deelnemers minder correcte antwoorden gaven voor moeilijkere taken en bevestigden dus de bevindingen van een groot aantal eerdere onderzoeken. Bovendien had taakcomplexiteit ook invloed op taakgedragsmetingen. Een toename in taakcomplexiteit van een eenvoudige feiten-zoekopdracht naar een meer complexe zoekopdracht leidde tot een toename van de tijd om de opdracht op te lossen en een afname van de browse-activiteit. Echter zoals uit de eerste studie bleek, is de relatie tussen taakcomplexiteit en verschillende metingen niet lineair. Wanneer een taak te complex wordt, wordt geen verdere toename in tijd waargenomen en neemt het aantal zoekopdrachten af. Hoewel deze twee studies de relatie tussen taakcomplexiteit en taakprestaties bevestigden, toonden ze ook aan dat er enkele nuances zijn in deze relatie.

Bovendien werd de invloed van het structureringstype bestudeerd in de hoofdstukken 3 en 4. Zoals in beide hoofdstukken is vermeld, was informatie traditioneel hiërarchisch georganiseerd, wat verschillende beperkingen heeft met betrekking tot flexibiliteit en vindbaarheid. De studies in hoofdstukken 3 en 4 vergeleken de hiërarchische en de tagging-aanpak op twee aspecten van het oplossen van informatieproblemen: het hergebruik van informatie (hoofdstuk 3) en het opslaan en organiseren van informatie (hoofdstuk 4). Beide onderzoeken hebben aangetoond dat geen van de twee systemen voor het organiseren van informatie als beter dan de ander kan worden beschouwd. Een nadeel van het tagging-systeem was dat de deelnemers er nauwelijks ervaring mee hadden. Hoewel mensen een voorkeur voor tags uiten wanneer er expliciet naar wordt gevraagd, vallen ze in hun daadwerkelijke gedrag terug op het gebruik van een hiërarchisch systeem. Het leren van iets nieuws vergt inspanning en dus vertrouwen ze op het systeem waar ze al aan gewend zijn. Niettemin presteerden de deelnemers aan beide onderzoeken die een tagging-systeem gebruikten niet slechter in vergelijking met deelnemers die een hiërarchisch systeem gebruikten. Er kan daarom worden verondersteld dat ervaring met tagging de tot nu toe getrokken conclusies zouden kunnen veranderen.

In de hoofdstukken 4 en 5 werd de invloed van het ondersteunen van studenten met een classificatieschema onderzocht. Voorkennis hebben over een onderwerp is een voordeel bij het oplossen van gerelateerde informatieproblemen. Voorkennis van experts is gestructureerd in goed ontwikkelde cognitieve schema's, die worden gebruikt om nieuwe informatie te vergelijken met wat al bekend is. Dientengevolge kunnen experts nieuwe informatie veel gemakkelijker verwerken dan beginners aangezien laatstgenoemden niet over zulke goed ontwikkelde interne kennisstructuren beschikken. Een van de meest belangrijke effecten in beide studies was dat deelnemers die ondersteuning kregen met een classificatieschema ook meer tags gebruikten die al aanwezig waren in hun eerdere vastgelegde kennisstructuren van de pre-test. Op basis van het classificatieschema konden deelnemers bekende concepten herkennen en die relateren aan de informatie die ze wilden opslaan en ordenen. Bovendien leidde de ondersteuning met een classificatieschema ertoe dat vaker gebruik werd gemaakt van classificaties en een meer gericht gebruik van classificatie-trefwoorden.

Dankwoord

Eindelijk, daar ligt die dan: mijn proefschrift. Het product van een lang traject van vallen en opstaan. Hoewel op dit moment van schrijven de allerlaatste hobbel (de verdediging) nog moet worden genomen, overheersen de gevoelens van blijdschap, opluchting en trots. Ik heb het toch maar mooi geflikt. Het was ook een leerzame tijd waarin ik onder andere veel geleerd heb over onderzoek doen, eye-tracking, statistiek, programmeren, informatievaardigheden en onderwijskunde én niet te vergeten over mezelf. Dit proefschrift staat dan weliswaar op mijn naam, maar zonder de hulp van anderen was dit niet tot stand gekomen. Hoog tijd dus om deze personen te bedanken.

Allereerst Saskia Brand-Gruwel en Halszka Jarodzka: jullie waren het beste begeleidingsteam dat ik me kon wensen. Saskia, onze samenwerking dateert al van voordat ik met mijn promotieonderzoek begon. In 2008 begonnen we aan het ontwikkelen van een cursus informatievaardigheden voor leraren. Ergens in 2009 kwam tijdens een treinreis voor dit project mijn ambitie om nog eens een keer te promoveren ter sprake. Aan het einde van de terugreis lag er een globaal idee dat we daarna verder hebben uitgewerkt in een onderzoeksplan. Dit karakteriseert ook jouw manier van begeleiden: gedreven, kritisch en oplossingsgericht met altijd oog voor het grote plaatje, maar ook altijd hartelijk en optimistisch. Al snel veranderde je rol van copromotor naar promotor en moesten we dus op zoek naar een nieuwe copromotor. Gezien de aard van mijn onderzoek met eye-tracking als belangrijkste dataverzamelingsmethodiek was het niet meer dan logisch om aan te kloppen bij Halszka. Halszka, ook jouw kritische begeleiding heeft mijn proefschrift beter gemaakt en mijn ontwikkeling als onderzoeker versterkt. Daarnaast heeft je natuurlijke vrolijkheid veel lastig denkwerk een stuk draaglijker gemaakt. De samenwerking met jullie beiden is dan ook iets waar ik met veel plezier op terugkijk. De sfeer was altijd goed. Onder het genot van een glas wijn aan het zwembad van het congreshotel op Cyprus een artikel bespreken met je promotor of de spanning van een goed gelukte presentatie van je afschudden met een spontane karaoke in de late uurtjes op een plein in Lund met je copromotor maakt het geheel toch een stuk dragelijker. Met het afronden van dit proefschrift is nu ook de laatste van de 3J's van Saskia klaar. Johan (2014) en Jimmy (2017) gingen mij al voor. Hoewel we geen van drieën over enige vorm van zangtalent beschikken, compenseren we dat hopelijk ruimschoots met de geleerde onderzoeksvaardigheden.

Zonder proefpersonen geen onderzoek. Daarom wil ik ook de scholen die mij de mogelijkheid hebben geboden om data bij hen te verzamelen hartelijk bedanken: het Bernardinuscollege te Heerlen, het Sophianum te Gulpen, het Connect College te Echt en het Porta Mosana College te Maastricht. Paola Oomen en Martin van Dijk wil ik bedanken voor het helpen verzamelen van de data. Jullie hulp heeft me veel tijd bespaard. Ook mijn collega's bij het Welten-instituut en het Ruud de Moor Centrum wil ik bedanken voor hun gezelligheid, steun en leerzame ervaringen. Dit geldt uiteraard voor alle (oud)aio's & postdocs: Andrea, Ángel, Annemarijn, Arnoud, Bernardo, Bettine, Bieke, Birgit, Charlotte, Da-

niël, Dirk, Elma, Eniko, Femke, Helen, Inge, Inne, Iwan, Jan, Janneke, Jérôme, Jimmy, Johan, Joyce, Kamakshi, Kiavash, Kim, Liesje, Ludo, Maren, Marieke, Marije, Martin, Michelle, Milou, Miriam, Paula, Pauline, Rory, Sandra, Sibren, en Thomas. Daarnaast heb ik ook veel geleerd en prettig samengewerkt met Karel, Marco en Peter in het MOOC-onderzoeksproject en met Marcel in het EVC-onderzoeksproject.

Buiten het werk heb ik ook altijd terug kunnen vallen op een fijne familie- en vriendenkring. Een aantal daarvan wil ik hier dan ook in het zonnetje zetten. Pa en ma, *nature* en *nurture* komen samen in jullie bijdrage aan dit proefschrift. De nieuwsgierigheid, eigenwijsheid en het optimisme dat ik van jullie heb geërfd in combinatie met het doorzettingsvermogen dat ik heb gekregen door jullie opvoeding en de steun die jullie hebben gegeven in het behalen van de hoogst mogelijke opleiding zijn onontbeerlijk geweest. Gerard, van je beste vrienden moet je het hebben. Ooit zijn we door onze ouders samen in één box gezet en daarna is het nooit meer goed gekomen. Gelukkig kan ik nu eindelijk antwoord geven op die éne prangende vraag van jou: "*Is het werkstukje voor de juf al af?*" Ja, nu wel. Ook mijn schoonouders wil ik hier niet onbenoemd laten. Jullie hebben me altijd gesteund en aangemoedigd om dit traject tot een goed einde te brengen. Gerard, wat had ik het je gegund als je dit nog mee had kunnen maken en wat zou je trots zijn geweest. Ik mis nog steeds onze discussies over sport en politiek en je welgemeende en goede adviezen. En dan Bart, jij bent de broer die het meest op mij lijkt. Het is daarom misschien niet vreemd dat jij ook aan een promotie-traject bent begonnen. Het was fijn om onze ervaringen te kunnen delen. Gedeelde smart is ten slotte halve smart. Met een stevige eindsprint ben ik je voorgebleven, maar ik heb er alle vertrouwen in dat jij ook spoedig op het podium staat. Verder heeft de R-community op StackOverflow en dan met name de 'GMT'-leden (o.a.: Ananda, Andrew, Arunkumar, Brodie, Catherine, David (2x), Edward, Florian, Frank, Jan, Michael, Pierre, Richard, Roland, Roman, Ronak, Soteris, Steven, Tamas, Tensibai, Tokhir, Uwe en Wouter) een belangrijke bijdrage geleverd aan mijn kennis van de statistische programmeertaal R. Dank voor alle inhoudelijke, technische en persoonlijke gesprekken.

Lest best. Mensen die mijn eetgewoontes kennen, weten dat ik altijd het beste voor het laatste bewaar. Zo ook hier. Pauline, jouw onvoorwaardelijke liefde en steun zijn van onschatbare waarde geweest gedurende de afgelopen jaren. Daarmee heb je me regelmatig door moeilijke tijden heen getrokken en zonder jouw bemiddeling had ik het laatste experiment helemaal niet uit kunnen voeren. Nu m'n proefschrift klaar is, heb ik weer zeeën vrije tijd beschikbaar. Dat we die samen op een leuke manier in gaan vullen, daar twijfel ik niet aan. Project 'Boris' kan nu tenminste echt van start! Je bent mijn lotje uit de loterij.

Luctor et emergo.


expertise
schema
tagging
bookmarks
learning
re-access
structuring
prior
support
complexity
hierarchical
behavior
gathering
storing
keywords
eye-tracking
information
cognitive
navigation
performance